AD_____

GRANT NUMBER: DAMD17-94-J-4270

TITLE: A New Generic Method for the Production of Protein-Based Inhibitors of Proteins Involved in Cancer Metastasis

PRINCIPAL INVESTIGATOR: Marshall H. Edgell, Ph.D.

CONTRACTING ORGANIZATION: University of North Carolina
at Chapel Hill
Chapel Hill, NC 27599-4100

REPORT DATE: August 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1996	Annual (1 Aug 95 - 31 Jul 96)

4. TITLE AND SUBTITLE
A New Generic Method for the Production of Protein-Based Inhibitors of Proteins Involved in Cancer Metastasis

5. FUNDING NUMBERS
DAMD17-94-J-4270

6. AUTHOR(S)
Marshall H. Edgell, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-4100

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

Our objective is to learn how to make inhibitors of proteins involved in metastasis. During 1995 we discovered that the protein framework (eglin c) we intended to use did not function in our assay system. During 1996 we constructed a circularly permuted form which does work. In parallel we have been using a truncated form (teglin) as a framework for inhibitor construction. Our first metastasis related target will be the proteinase stromelysin, but in the absence of a source of that protein have been using the proteinase papain. We have shown that teglin can be converted to a weak binder to papain. This is a milestone in that it documents that our presumption that we would be able to select binders for heterologous targets is correct. We have purified active stromelysin from small scale lysates of clones made in 1995 and are currently getting about ~1 mg of prostromelysin per liter of culture which if efficiently converted to active stromelysin will be sufficient for phage display assays. We have carried out thermodynamic measurements on our basic framework structure, eglin c. Biophysical measurements will be used in constructing high affinity inhibitors and these experiments provide baseline data to support that approach.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	36
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

MHE In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).
(no animals used this year)

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

MH Ebell

PI - Signature Date

Table of Contents

INTRODUCTION

Our objectives are to learn how to efficiently build proteins which will act as inhibitors to proteins involved in cancer metastasis. It is our expectation that such proteins can be constructed so as to be very specific for the desired target. While such inhibitors might be useful themselves as therapeutic molecules, they will certainly be useful as probes to define the issues associated with inactivating the target proteins; both the primary effects and side-effects.

The approach to be used in this project is to combine molecular genetics and protein biophysics to redirect to the target of interest the activity of a pre-existing protein which will serve as a framework onto which to mount the desired modifications. Molecular genetics will be used to extend the reach of traditional protein engineering. The idea is to make large libraries of structural variants and then use genetic screening and selection strategies to find the best performers. Traditional protein biophysics will then be used to explore the various classes of variants and to make models for what is leading to inhibition. This information will then be used in subsequent cycles of design, construction and screening.

The development cycle that we will employ to modify the wild-type eglin c into a new inhibitor is:

1. make our best design guess as to what changes will increase binding to the new target
2. construct a 'halo' of variants (~10^7) around the design
3. screen the variant library using phage display to find the best binders to the new target
4. characterize the binding classes using biophysical techniques (NMR, CD, ANS binding, etc.)
5. use the biophysical information and modeling to build hypotheses concerning binding
6. if affinities not high enough go to step 1

The simplest inhibitors bind to their targets close to or at the active site and interfere with activity simply by getting in the way and not 'letting go' of the target. That is, such inhibitors lower the energy of the Michaelis complex sufficiently that very few molecules reach the transition state. These inhibitors can be designed to poke a projection into a groove in the target or enfold a pocket over a projection on the target. As our initial protein for protein engineering we wanted a molecule that was small, well mannered and for which we had some reason to think might be structurally compatible with our first set of targets. Our choice was a small proteinase inhibitor, eglin c. This protein is exceptionally stable, has no disulfide bonds, is well characterized and binds very tightly to proteins similar to our first target which is stromelysin, a proteinase implicated in metastasis. Eglin c inhibits its normal targets, serine proteinases, by binding so tightly in the Michaelis complex, that the protein cannot be raised into the transition state. A ten amino acid loop in eglin c binds within the active site groove of the native serine proteinase targets.

High affinity binding requires a sequence (binding epitope) which is compatible with the target and a set of structural constraints on that sequence which prevent it from spending much time in non-productive conformations. The engineering task, which we have set ourselves, is to replace the wild-type binding epitope with one suitable for the target and then to construct a set of new constraints to move the binding epitope into the high affinity domain. Our initial target is stromelysin and hence an appropriate binding epitope is already known, that is, a substrate sequence preferred by the proteinase. Building an inhibitor then reduces to finding a suitable series of structural constraints that can be imposed by the eglin framework on the new binding epitope.

We are pursuing two sub-lines of investigation. One is to expand our information about eglin c as a suitable framework for protein engineering and the other is to start the protein engineering with what we already know.

BODY

During this second year we have been putting together the components necessary for the engineering activities and testing materials made during the first year. We have:

1. constructed an eglin c variant suitable for phage display
   a. constructed a circularly permuted form of eglin c to use in phage display experiments since phage displaying wt eglin do not bind to the appropriate protein (subtilisin).
   b. showed that expression levels of peglin (permuted eglin) in *E. coli* approximates that of eglin c. This implies that the changes in the molecule have not unduly destabilized peglin .
   c. showed that phage displaying peglin bind to the cognate protein (subtilisin) and give a 200-fold enrichment
2. shown that a truncated form of eglin c can be altered to bind to the proteinase papain as a model for making inhibitors for proteins involved in metastasis
   a. a truncated form of eglin (18 amino acid closed loop) with the subtilisin substrate sequence replaced by the papain substrate sequence binds to papain and gives 30-fold enrichment in phage display
   b. made a 'white' version of the papain binder for use in the next round of testing
   c. made a 'binding epitope' phage display library
   d. selected six variants that bind to papain better than the original construction.
3. continued efforts to make purified stromelysin for phage binding.
   a. verified our activity test using a gift of active stromelysin
   b. verified our prostromelysin activation protocol using a gift of prostromelysin.
   c. showed that two clones made during first year do indeed express a prostromelysin from which active stromelysin can be produced.
   d. purified his-tagged prostromelysin on a small scale using nickel columns. From this experiment we estimate that several liters will produce sufficient stromelysin for phage display experiments.
   e. started large scale purification (two liter lysates)
4. carried out a series of baseline thermodynamic characterizations of eglin c.

1. Construction of an eglin c variant suitable for phage display

Our basic approach is to extend the reach of traditional protein engineering by constructing large libraries of structural variants around a central design concept and then using the phage display system to screen the population for binders. Hence it came as a considerable shock when we discovered that the framework protein on which we intended to do all of our engineering, wild-type eglin c, when fused to the M13 gene III protein in the M13 particle, does not bind to a target to which the free inhibitor normally binds (e.g. subtilisin). We presume that the problem is that when the phage are attached to eglin c via it's C-terminus it blocks access to the eglin c binding epitope (see Figure 1). That is, the eglin c C-terminus appears to be too close to the loop containing the residues that bind with the target proteinase.

Last year, to try to move the gene III fusion point further away from the binding epitope, we made eglin c variants with various C-terminal truncations and made the M13 gene III fusions via linkers of 5 and 9 prolines. None of these eglin c variants bound to subtilisin.

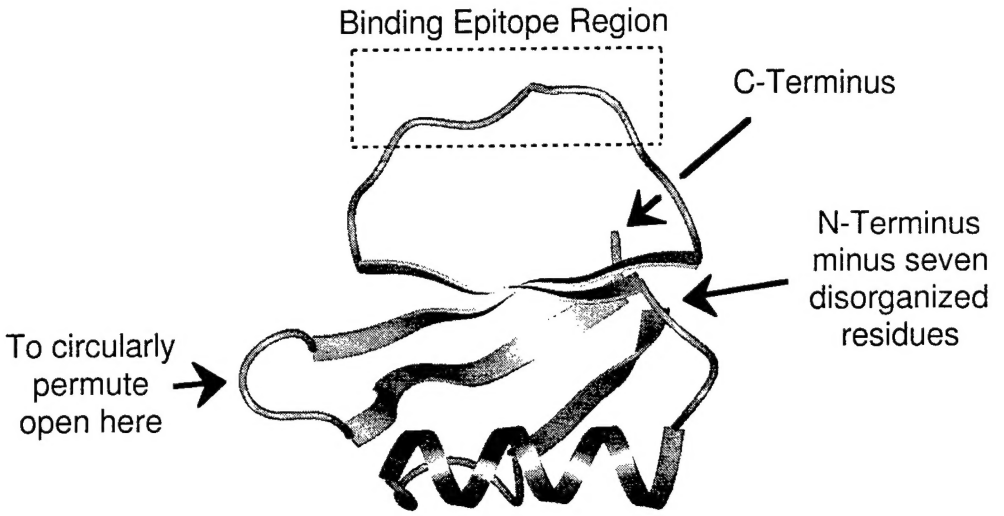


Figure 1. Ribbon Diagram of Eglin C. Note that attaching a phage particle to the C-terminus of eglin c might block access to the binding epitope. Our construction of a circularly permuted eglin removes the seven disorganized residues from the N-terminus, adds a four residue tight turn to connect the N- and C-terminal ends, and opens the protein to create new N- and C-termini at the point indicated in the figure.

1a. Construction of a circularly permuted version of eglin called peglin.

Last year we also constructed a circularly permuted version of eglin c in which the C-terminus has been moved to the side of the protein opposite from the residues that bind with the target. We designed an eglin variant in which the wild-type N and C termini were joined together and new termini were created by opening up a tight turn on the opposite side of the protein (Figure 2).

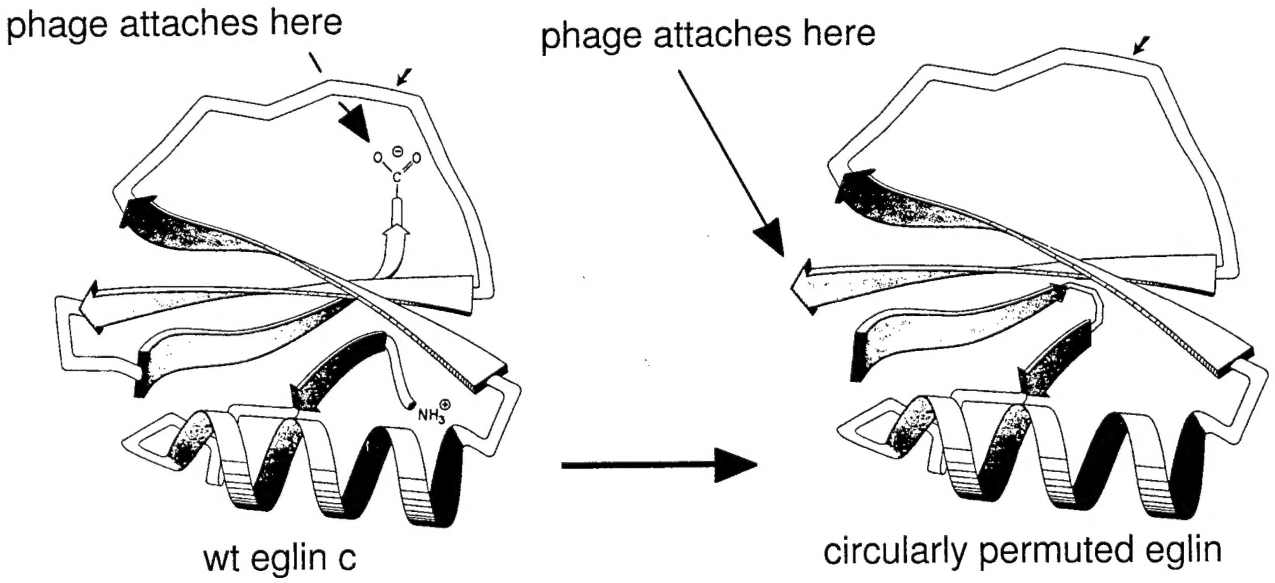


Figure 2. Circularly Permuted Version of Eglin C. The circularly permuted version of eglin, which we have called peglin, provides a phage attachment site much farther away from the target protein binding site present in the large (ten amino acid) loop on the 'top' of the molecule.

1b. Expression levels of peglin (permuted eglin) in *E. coli* approximates that of eglin c. This implies that the fold is not massively destabilized by this structural modification.

The peglin construct is in the pET28 (Novagen, Madison, WS) expression vector which is driven by a T7 promoter. The plasmid was transformed into a BLRLysS *E. coli* cell line (Novagen, Madison, WS) and grown in 2YT plus chloramphenicol (34 ug/ml) and kanamycin (50 ug/ml) to an OD_{600} of 0.55. The cells were then pelleted and frozen with dry ice/alcohol. For harvesting pellets were thawed in 1/10 original volume of 50 mM tris (pH 8.0), 2 mM EDTA, and lysozyme added to 100 ug/ml. After 30 minutes at room temperature the cells were sonicated for 2 minutes at full power. The lysate was cleared by centrifugation at 14,000 rpm for 30 minutes. Samples boiled for 10 minutes in 1% SDS were examined on 15% polyacrylamide gets, 1% SDS. What one sees on these gels is that there is at least as much peglin relative to the coli proteins (lanes 1 through 6) in the lysates as there is eglin (lane 7).

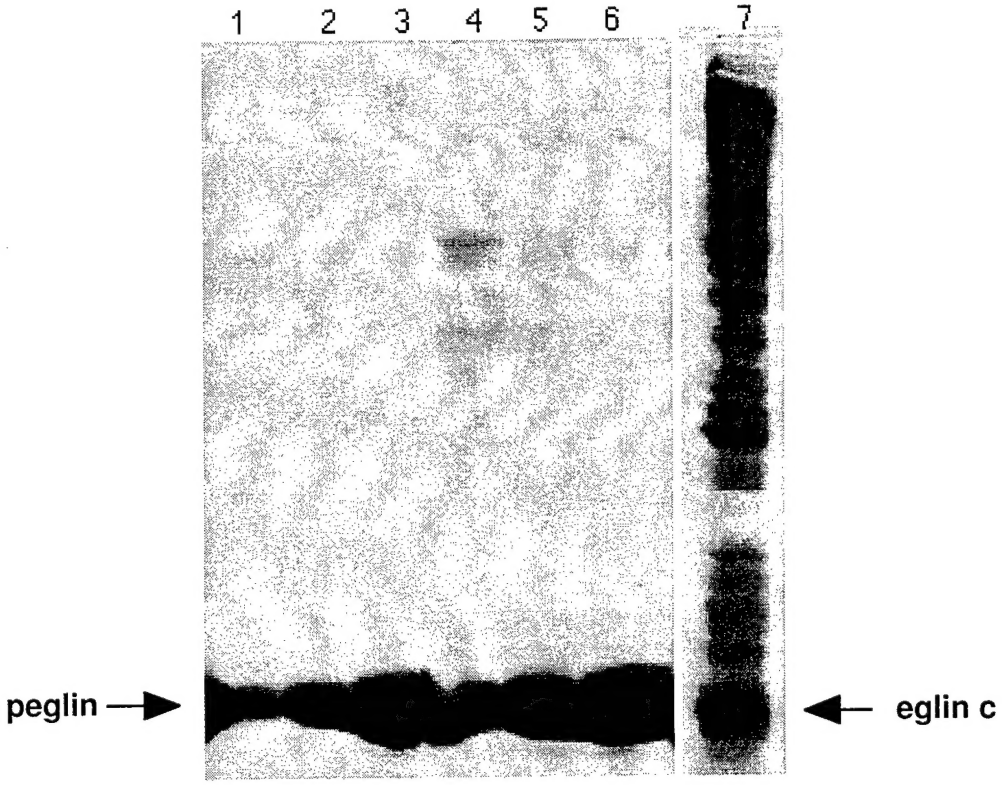


Figure 3. *E.coli* Lysates expressing peglin or eglin c.
Lane 1. Extract from Nova Blue DE3 with PEGLIN 0 hr after induction.
Lane 2. Extract from Nova Blue DE3 with PEGLIN 1.5 hr after induction.
Lane 3. Extract from Nova Blue DE3 with PEGLIN 3.0 hr after induction.
Lane 4. Extract from BLR with PEGLIN 0 hr after induction.
Lane 5. Extract from BLR with PEGLIN 1.5 hr after induction.
Lane 6. Extract from BLR with PEGLIN 3.0 hr after induction.
Lane 7. Extract from BLR with EGLIN C 0 hr after induction.
Note that there appears to be more peglin relative to the coli proteins in the lysates than eglin.

1c. Peglin binds to its cognate protein target (subtilisin) and gives a 200-fold enrichment.

We have now shown that this permuted version of eglin c, called peglin, does indeed function in the phage display system. This was done by moving the peglin construction into M13 mBAX (Brian Kay, personal communication) so that the peglin protein sequence is fused with the phage gene III protein. 96 well plates were coated with the cognate target protein, subtilisin, by adding subtilisin to wells of Corning

ELISA flat bottom 96-well plates (#25801) at a concentration of 25 ug/ml and incubating at 4C overnight. Any remaining protein binding sites were then blocked with 1% BSA solution for 12 hour at room temperature. To test for binding we use a blue/white screen to detect enrichment of the putative binder phage. Hence our peglin containing phage also carries the alpha fragment gene to provide blue plaques. For this purpose we have constructed an alpha fragment mutant of the basal phage display phage, mBAX, which gives white plagues. A mixture of 1:10 putative binding phage (peglin in blue mBAX) to non-binding phage (mBAX white) was added to the subtilisin coated wells and incubated at room temperature for 4 hours. Wells were washed eight times with PBS-0.1% Tween 20 and then phage were eluted from the plate using 150 ul 50 mM glycine-HCl (pH 2.0). The solution containing eluted phage was neutralized with 150 ul 200 mM $NaPO_4$ (pH 7.5). The titer of phage was determined on plates containing x-gal and IPTG. Two isolates of peglin were tested. While the isolates were expected to be identical peglin 2 consistently binds better than peglin 1 (Table 1). Sequence analysis is currently being carried out to determine the actual sequence of the peglin 2 isolate.

SAMPLE	Input Phage Concentration pfu/ml	Input Ratio binders:non-binders	Output Ratio binders:non-binders	Enrichment
peglin 1 (round 1)	$2.6x10^9$	>1:25	>1:1	>25
peglin 1 (round 2)	$2.4x10^7$	1:7	>1:1	>7
peglin 2 (round 1)	$8.4x10^{10}$	1:167	>2:1	>334
peglin 2 (round 2)	$1.9x10^8$	>1:200	>2.3:1	>466

Table 1. Peglin Binds to Subtilisin as Indicated in a Phage Display Assay.

2. An Alternative Framework for Binding: Teglin

We have been using a truncated form of eglin c, which we are calling teglin, as an alternative framework. Leatherbarrow (1991) showed that an eighteen amino acid circular peptide consisting of the amino acids in the eglin c loop (Figure 4) containing the binding epitope and the amino acids from the

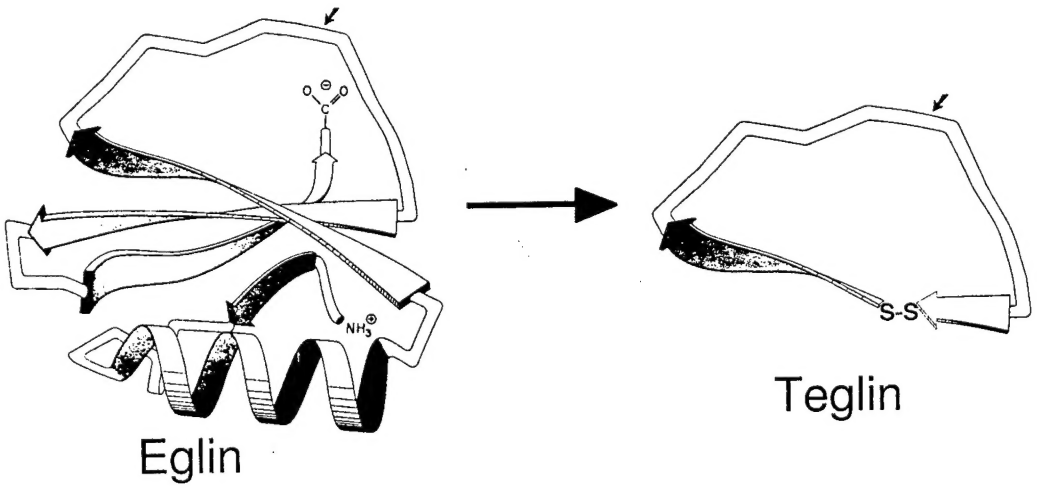


Figure 4. Conversion of eglin to teglin. The ten amino acid loop containing the binding epitope and the underlying beta strands containing two framework to loop salt bridges is removed and joined via a cysteine bond. The teglin structure would, of course, not maintain the beta sheet residues from eglin in a beta sheet conformation in teglin.

underlying strands of beta sheet closed together via a cysteine bond was fully as active as native eglin c. We have determined that phage displaying this sequence bind very tightly to subtilisin.

2a. Truncated form of elgin can be altered to bind to papain.

We have replaced the binding epitope specific for subtilisin with the substrate peptide specific for the proteinase papain Figure 5). Papain is being used here as a model for a metastasis related protein until

```
teglin    CysGlySerProVal ThrLeuAspLeu ArgTyrAsnArgValArgValPheCys
teglinp   CysGlySerProVal PheAlaGluAla ArgTyrAsnArgThrArgValPheCys
```

Figure 5. Papain Version of Teglin. The binding site of teglin (shown in bold) is removed and replaced by the cleavage target of stromelysin.

we prepare enough stromelysin, our first choice target, for binding studies. We have now shown that the truncated variant binds weakly (8 to 37-fold enrichment in phage display) to papain. This represents a break-through in the sense that this experiment documents the central thesis of our project, a capacity to retarget a naturally occurring inhibitor to a heterologous target.

Input Phage Mixture	Input Ratio (B:W)	Average Input Ratio (B:W)		Output Ratio (B:W)	Average Output Ratio (B:W)		Fold Enrichment (W)
Teglin Ø (W) + m663 (B) (active papain)	45:1 235:1	140:1		29:1	29:1		4.8
Teglin (B) + m663 (W) (inactive papain)	18:1 68:1	42:1		13:1 11:1	12:1		3.5
Papain Ø 13a (W) + m663 (B) (active papain)	5:1 8:1 4.9:1	6:1		.8:1 .5:1	.7:1		9.2
Papain Ø 13a (W) + m663 (B) (inactive papain)	(22:1) 4:1 8:1	(11:1) 6:1		.3:1 .3:1	.3:1		(37) 20
Papain Ø 13b (W) + m663 (B) (active papain)	1.6:1 4:1 1.1:1	2.2:1		.3:1 .2:1	.25:1		8.8
Papain Ø 13b (W) + m663 (B) (inactive papain)	1.4:1 9:1 1.3:1	2.3:1		.18:1 .15:1	.17:1		14
m666 (W) + m663 (B) (active papain)	1.3:1 .5:1 1.2:1	1:1		.9:1 .7:1	.8:1		1.2
m666 (W) + m663 (B) (inactive papain)	.9:1 1.9:1	1.4:1		.7:1 1.8:1	1.3:1		1.1

Table 2. Phage Binding to Papain. Phage mixtures were assayed in 96-well plates in wells coated with papain as per the procedures described in section 1c. The papain was active or inactivated with EDTA as indicated. Inactivated papain is expected to have the native confirmation but to be unable to cleave weakly binding phage. Numerical values in parentheses represent figures which are estimates due to the low number of input (white) phage.

2b. Construction of a white variant of the papain binder for use in next round of phage display.

We wish to carry out several cycles of binding improvement. To optimize our capacity to find these variants we want to mix library variants that generate blue plaques with the starting construction which has been designed to give white plaques. To do that we wanted to construct a second phage which gave white plaques and was identical in growth rate to our blue plaque phage. To do that we used PCR to amplify the entire mBAX plasmid minus a ~200 bp region within the beta galactosidase al;pha fragment (Figure 6.).

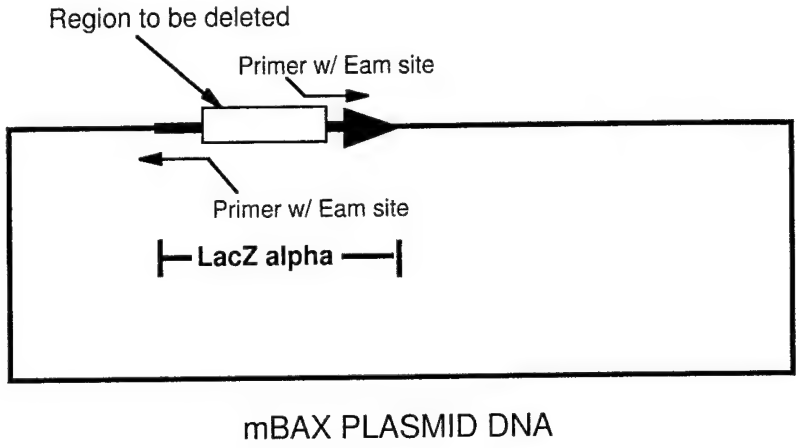


Figure 6. Construction of a 'White' mBAX Phage. A ~200 bp deletion in mBAX was created by using PCR and primers containing EAM restriction sites. Restriction enzymes recognizing these sites cut outside of the site allowing one to cut and splice anywhere one likes within a sequence independent of sites inside the target DNA. Circular RF DNA was used as the template for the PCR.

2c. Construction of the first papain binder library.

We have constructed a small library of variants (~10^4) which randomized several of the amino acids in the binding epitope or near the binding epitope (Figure 7). This library was screened using phage display protocols described in 1c. but against papain.

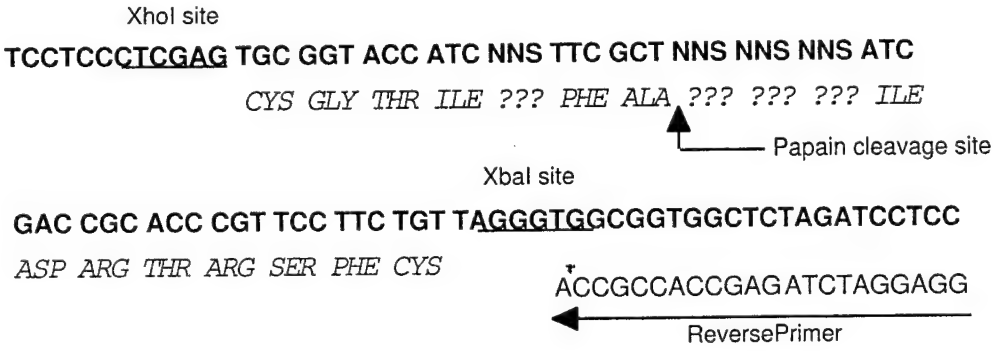


Figure 7. Construction of Library to Improve Papain Binding Site. Four residues were randomized around the putative site of papain binding which is based on the sequences preferred for cleavage by papain.

2d. Selection of better binding variants from a library.

We amplified the library and then subjected the population to four rounds of panning to find variants which would bind to papain better than the starting phage. Six variants were selected which bound

better than the original construction (Table 3). We are in the process of analyzing these variants to determine the sequences of the improved binders.

Clone Number	Input Conc	Input Ratio (B:W)	Output Conc	Output Ratio (B:W)	$Output_{ratio}$ divided by $Input_{ratio}$
L1	$3.5x10^{11}$	.9:1	$4.1x10^{6}$	1.2:1	1.3
L2	$2.7x10^{11}$	1.5:1	$1.8x10^{6}$	.3:1	0.2
L3	$5.3x10^{11}$	4.3:1	$1.8x10^{6}$	.3:1	0.1
L4	$2.4x10^{11}$	5:1	$2.6x10^{6}$	1.6:1	0.3
L5	$5.2x10^{11}$	3:1	$2.6x10^{6}$	.2:1	0.1
L6	$5.4x10^{11}$	3.5:1	$2.5x10^{6}$	.3:1	0.1
L7	$5.6x10^{11}$	4.6:1	$1.3x10^{6}$	.4:1	0.1
L8	$3.4x10^{11}$	1.1:1	$2.8x10^{6}$	1.5:1	1.4
L9	$8.1x10^{11}$	2.1:1	$2.7x10^{6}$	.3:1	0.1
L10	$8.9x10^{11}$	3.5:1	$3.1x10^{6}$	.3:1	0.1
L11	$2.5x10^{11}$	1.8:1	$2.5x10^{6}$	1.7:1	0.9
L12	$6.5x10^{12}$	4.5:1	$1.3x10^{6}$	.3:1	0.1
L13	$5.6x10^{11}$	1.9:1	$4.2x10^{7}$	111:1	58.0
L14	$3.1x10^{11}$	1.8:1	$2.5x10^{7}$	81:1	45.0
L15	$8.0x10^{12}$	6.5:1	$1.2x10^{6}$	.35:1	0.05
L16	$4.5x10^{11}$	4.6:1	$1.9x10^{7}$	11:1	2.4
L17	$9.2x10^{11}$	3:1	$2.7x10^{6}$	.3:1	0.1
L18	$4.9x10^{11}$	2.5:1	$1.3x10^{6}$	.4:1	0.1
L19	$1.4x10^{11}$	.4:1	$1.6x10^{6}$	.4:1	1.0
L20	$7.2x10^{11}$	5.5:1	$1.5x10^{6}$	.2:1	0.04
L21	$1x10^{12}$	2.6:1	$2.5x10^{6}$	.4:1	0.2
L22	$7.8x10^{11}$	2.9:1	$2.0x10^{6}$	.5:1	0.2
L23	$4.5x10^{11}$	6.5:1	$1.2x10^{6}$	.3:1	0.05
L24	$4.8x10^{11}$	2.7:1	$.9x10^{6}$	.6:1	0.2
...	...	...	...	...	...
L34	$7.1x10^{10}$	10:1	$1.5x10^{7}$	21:1	2.1

Table 3. Phage Display Binding Data for Individual Clones. Individual clones from the library shown in figure 7 were picked after four rounds of panning and tested compared for binding with the white version of our initial papain binding phage.

3. Stromelysin Production

Our objective is to learn how to construct protein-based inhibitors to proteins involved in the metastasis of cancer. As a model system we have chosen the proteinase stromelysin since it has been strongly implicated in metastasis and is a proteinase and so must be able to accommodate a peptide chain within its active site. However, there is no commercial source for stromelysin so we are currently involved in constructing an expression vector containing the stromelysin sequence. A complicating feature of the construction is that while the mature sequence for stromelysin is known, it was less clear what additional sequences (prepro?) are necessary for maturation of the enzyme in *E. coli.* at the time we designed our clones. (It is now known that the mature form of stromelysin can be made in *E. coli* without toxic effects and with full activity.) We have constructed five variants and each produces a stable protein in *E. coli.*

3a. Verified an activity test for stromelysin using a gift of active stromelysin.

Dr. Paul Cannon at Roche Biosciences in Palo Alto, CA provided us with a sample of an active mature form of human stromelysin, originally from Agouron Pharmachutical, which is 18 kD and produced in *E. coli*. This material was diluted in our laboratory to a working concentration of 1 mg/ml in 25 mM tris (pH 7.25), 10 mM $CaCl_2$, and 0.05% Brij-35 and stored at -70C in small aliquots. For our stromelysin assay we used an activity assay designed for vertebrate collagenase (Weingarten et al., 1985) utilizing the hydrolysis of a thiopeptide substrate Ac-pro-leu-gly-[2-mercapto-4-methyl-pentanoyl]-leu-gly-OE_t, purchased from Bachem Biosciences (Philadelphia, PA) and Ellman's reagent (DTNB or 5,5'-dithio-bis(2-nitrobenzoic acid)) purchased from Sigma Biochemical Co. (St. Louis, MO). The assay conditions are 50 mM MES (pH 6.0), 10 mM $CaCl_2$, 106 mM thiopeptide substrate and 1 mM DTNB in a 100 ul final volume (Qi-Zhuang, et al., 1992). Thiopeptide is prepared at 76 mM in 80% acetic acid and is stable at -20C for several months. DTNB solution is prepared at 20 mM in 95% EtOH and is stable for several months at -20C. MES is prepared as a 1 M solution, filter sterilized and stored at -20C in 10 ml aliquots. $CaCl_2$ is prepared as a 1 M solution and autoclaved. A thermomax plate reader was used to monitor activity as indicated by change in adsorbance at 405 nM at room temperature (~24C) for 30 minutes with 11 second intervals between readings. A standard curve for the Agouron stromelysin plotting V_{max} versus stromelysin concentration is shown in Figure 8.

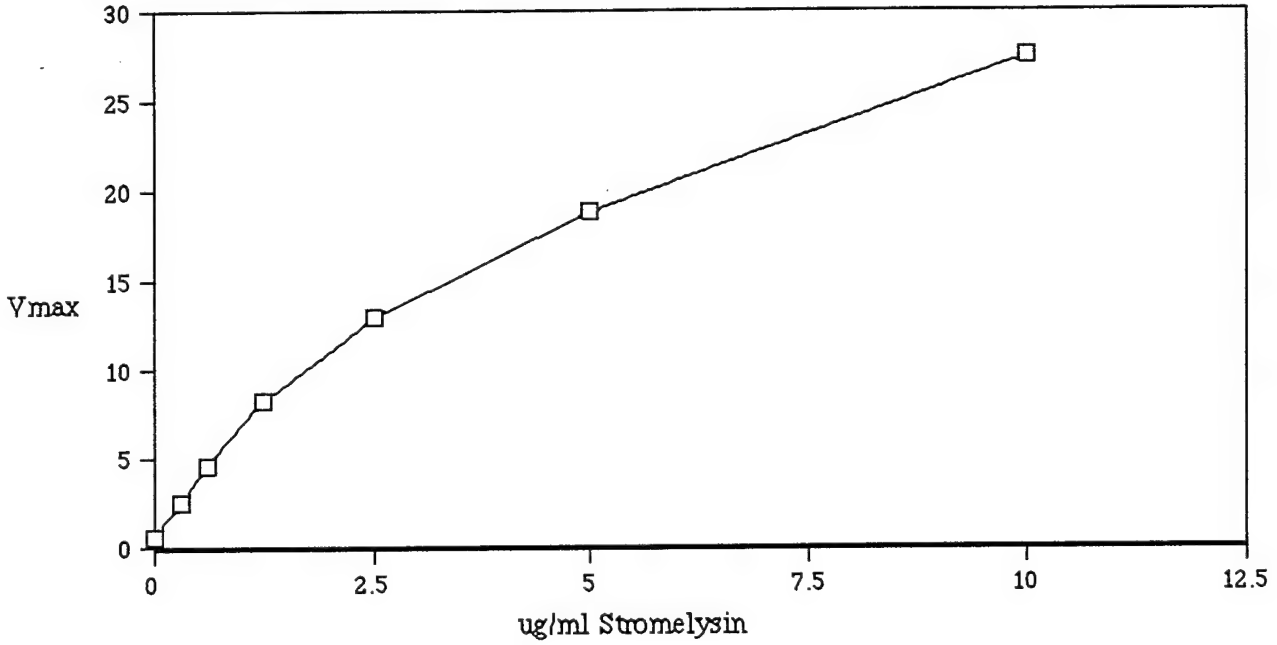


Figure 8. Standard Curve for Stromelysin. V_{max} versus stromelysin concentration.

3b. Verified that we can activate prostromelysin.

Dr. Paul Cannon also provided us with a sample of full length human prostromelysin-1 purified from IL-1 stimulated human gingival fibroblast conditioned medium. The material is stored at a concentration of 0.8 mg/ml in 50 mM tris (pH 7.4), 0.2 NaCl, 5 mM $CaCl_2$, and 0.02% NaN_3 at -70C in 100 ul aliquots. The trypsin method (Marcy et al., 1991) was tested for activation using the assay described in 3a to measure resultant stromelysin activity. Trypsin processes the 58 kD prostromelysin to a 45 kD active form which partially autoprocesses to smaller active forms around 28 kD. Trypsin does not

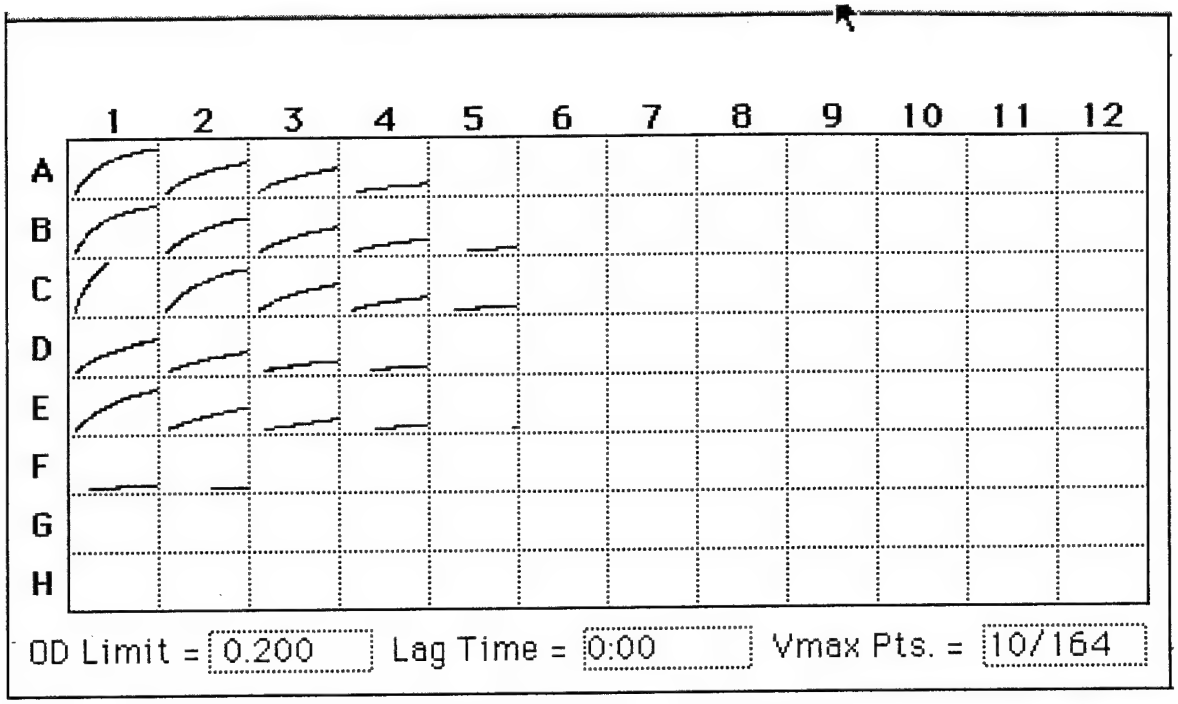


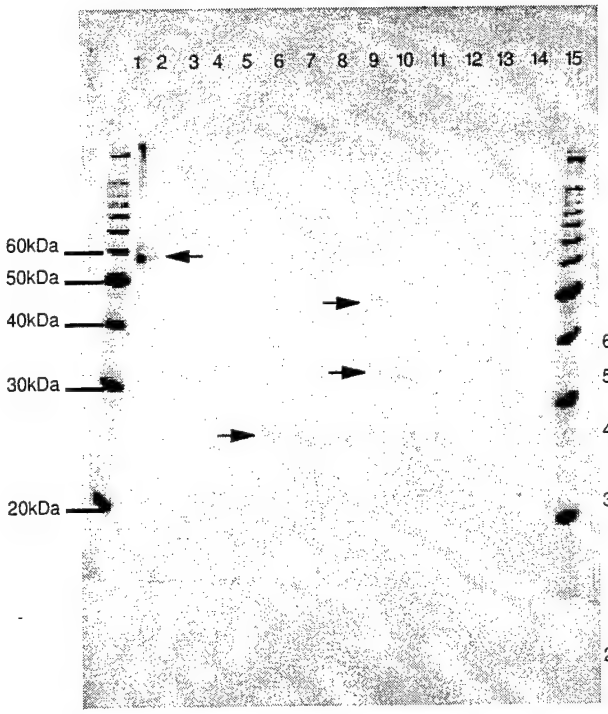
Figure 9. Trypsin Activation of Prostromelysin. Samples of mature stromelysin (short form) or trypsin treated prostromelysin. Reaction conditions were: 50 mM Tris (pH 7.4), 5 mM $CaCl_2$ and 0.2 M NaCl with incubation for 30 minutes at 37C. (Trypsin by does not itself convert substrate to color.)

Rows A, B, C. stromelysin: 5 ug/ml , 2.5 ug/ml, 1.25 ug/ml, 0.625 ug/ml

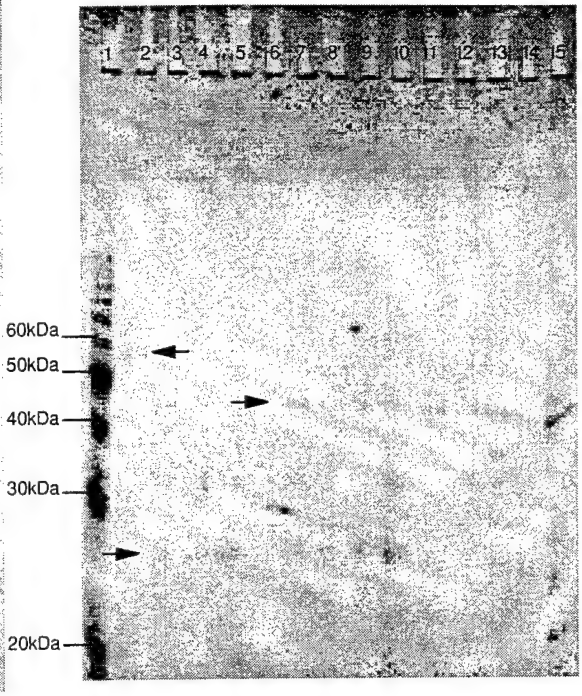
Rows D,E. prostromelysin 100 ug/ml, 50 ug/ml, 25 ug/ml, 12.5 ug/ml; each contains 25 uM trypsin

Row F. same as D & E plus 75 uM trypsin inhibitor

act on the substrate to produce color (data not shown). Approximately 2-3% of the prostromelysin was activated in this experiment (Figure 9). Longer times did not increase the yield (data not shown). A wider range of trypsin concentrations was tried using a gel assay to monitor conversion to a shorter form since this assay uses much less material (Figure 10). Prostromelysin was treated with 100 uM, 25 uM, and 4.16 uM trypsin and assayed for production of 48 and 28 kD forms of stromelysin. Treatment with 25 uM trypsin lead to conversion within 10 minutes. 45 minutes of treatment with trypsin at any of the concentrations tested lead to complete degradation of the stromelysin. The best activation treatment within the range tested is 4.16 uM trypsin for 10 minutes. We quantitated the amount of activity generated by this procedure using the activity test described in 3a. This procedure converted approximately 9% of the prostromelysin material to active stromelysin (Figure 11).



1 10kDa Marker
2 Prostromelysin
3 Prostrom. 10min. 100μM trypsin
4 Prostrom. 20min. 100μM trypsin
5 Prostrom. 30min. 100μM trypsin
6 Prostrom. 10min. 25μM trypsin
7 Prostrom. 20min. 25μM trypsin
8 Prostrom. 30min. 25μM trypsin
9 Prostrom. 10min. 4.16μM trypsin
10 Prostrom. 20min. 4.16μM trypsin
11 Prostrom. 30min. 4.16μM trypsin
12 Prostrom. 45min. 100μM trypsin
13 Prostrom. 45min. 25μM trypsin
14 Prostrom. 45min. 4.16μM trypsin
15 10kDa Marker



1 10kDa Marker
2 Prostromelysin
3 Prostrom. 10min. 4.16μM trypsin
4 Prostrom. 20min. 4.16μM trypsin
5 Prostrom. 30min. 4.16μM trypsin
6 Prostrom. 10min. 4.16μM trypsin +PMSF
7 Prostrom. 10min. 2μM trypsin
8 Prostrom. 20min. 2μM trypsin
9 Prostrom. 30min. 2μM trypsin
10 Prostrom. 60min. 2μM trypsin
11 Prostrom. 10min. 1μM trypsin
12 Prostrom. 20min. 1μM trypsin
13 Prostrom. 30min. 1μM trypsin
14 Prostrom. 50min. 1μM trypsin
15 Prostrom. 60min. 1μM trypsin

Figure 10. Conversion of Prostromelysin to Stromelysin. (P) prostromelysin at 58 kD. (S) 45 kD stromylysin. Unlettered arrows indicate the 28 kD from of stromelysin. Digestions were done in 50 mM Tris (pH 7.4), 5 mM $CaCl_2$, 0.2 M NaCl at 37C.

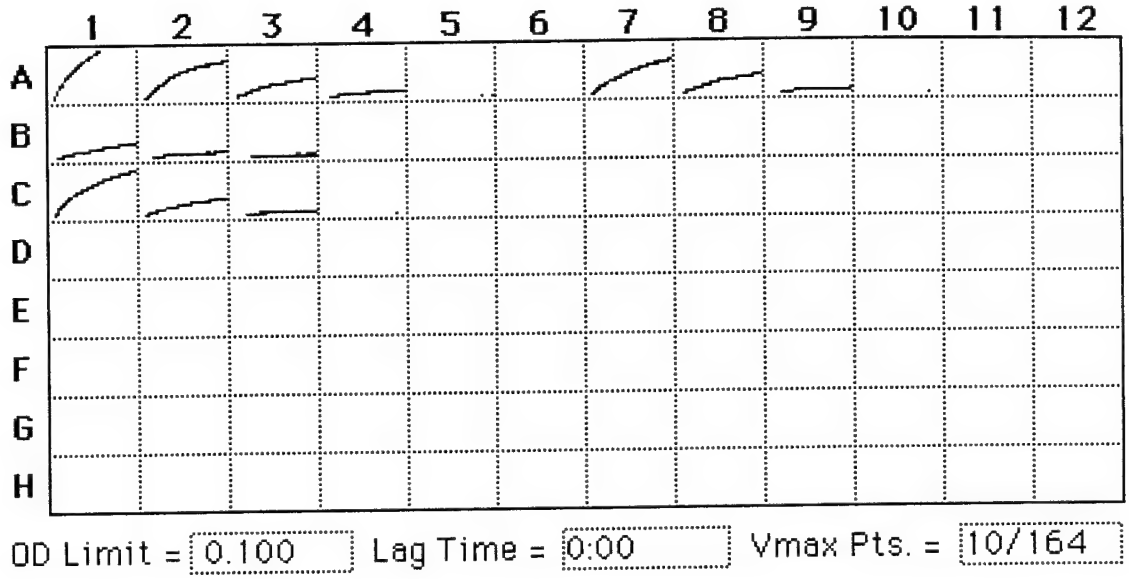


Figure 11. Trypsin Activation of Prostromelysin at 4.16 uM Trypsin. Colorimetric assay for stromelysin. Each box represents a time course of color development for one well in a 96-well plate. The substrate is Ac-Pro-Leu-Gly-[2-mercapto-4-methyl-pentanoyl]-Leu-Gly-OEt. Reactions were at room temperature for 30 minutes with readings every 11 seconds. Measurements were taken at 405 nm. The OD represented by the box height is 0.1.

Row A. Mature form of stromelysin (from Agouron)
Wells 1-5: 5ug/ml, 2.5 ug/ml, 1.25 ug/ml, 0.625 ug/ml, 0.31 ug/ml
Well 6: blank
Wells 7-11: 5ug/ml, 2.5 ug/ml, 1.25 ug/ml, 0.625 ug/ml, 0.31 ug/ml

Row B. Prostromelysin (from Agouron)
Wells 1-6: 48 ug/ml, 24 ug/ml, 12 ug/ml, 6 ug/ml, 3 ug/ml, 1.5 ug/ml, 0.75 ug/ml

Row C. Prostromelysin (from Agouron) treated for 30 minutes at 37C with 4.16 uM trypsin.
Wells 1-6: 48 ug/ml, 24 ug/ml, 12 ug/ml, 6 ug/ml, 3 ug/ml, 1.5 ug/ml, 0.75 ug/ml

3c. The two putative prostromelysin clones produced during the first year of this project do indeed produce activatible prostromelysin.

Last year clones for prostromelysin were constructed using PCR to amplify the desired sequence from a stromelysin 1 non-expression clone DNA provided by Dr. Lynn Matrycian. The amplified sequences were cloned in either pET 3d or pET 28a (Novagen, Inc. Madison, WS) downstream of a T7 promoter (pET 3d) or a T7lac promoter (pET 28a). These promoters are not transcribed in bacteria which do not have a source of the T7 RNA polymerase. Constructions are maintained in hosts without the T7 polymerase since high levels of expression are usually growth inhibitory and hence lead to growth advantages for variants which have lost either the protein gene sequences or the expression machinery. Our plasmid constructs were transformed into BLR (DE3) pLysS, a host expressing T7 RNA polymerase under tight control. Five independent transformation isolates were tested for activatible prostromelysin. Cultures were grown in LB at 37C with shaking at 250 rpm to an OD_{600} between 0.6 and 0.9. Expression was induced by adding IPTG to 0.4 mM for the 3d hosts and 1.0 mM for the 28a hosts and the cultures were then grown for an additional 3 hours. The cultures were chilled on ice for 5 minutes and the cells collected by pelleting at 4C. The pellets were washed in 0.25 growth volumes of 50 mM Tris (pH 8.0), 10 mM $CaCl_2$. The cells were pelleted again and frozen at -70C. To lyse the cells the pellets were thawed in a water/ice bath, resuspended in cold 1/10 growth volume 50 mM Tris (pH 8.0), 10 mM $CaCl_2$ and lysozyme was added to 0.2 mg/ml. After 20 minutes on ice the material was sonicated at 90% intermediate output (Branson Sonifier) for 1 minute. This lysate was cleared by centrifugation in an

Eppendorf microcentrifuge. The supernatant was collected and stored at 4C. Total protein was determined.

To test for activity the crude lysates were treated with trypsin to activate the prostromelysin. The resultant solution was then tested for stromelysin activity as per 3a (Figure 12).

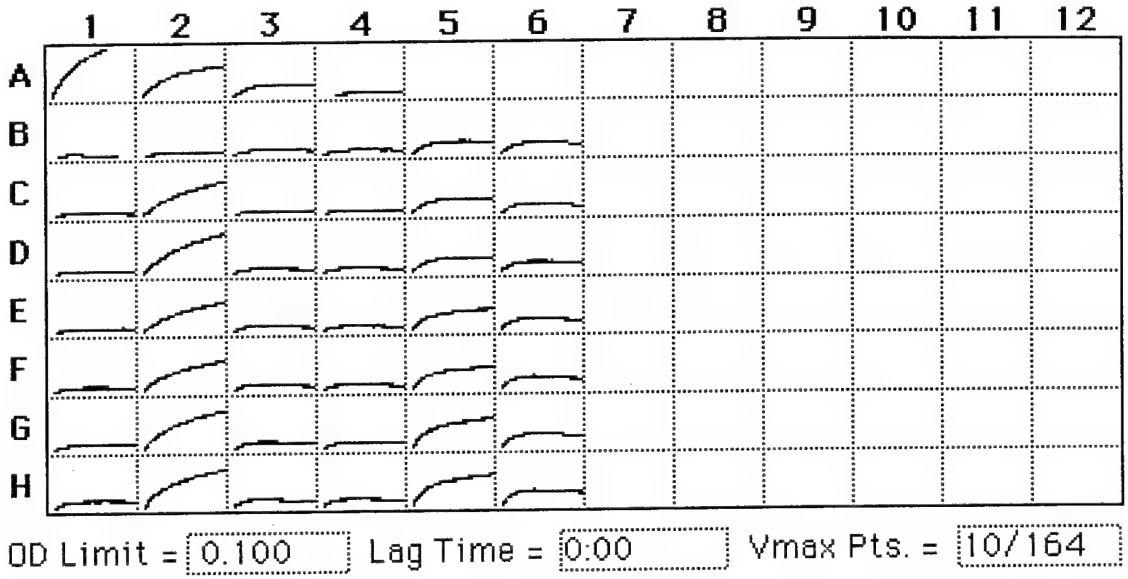


Figure 12. Activation of Prostromelysin (our clones) in Crude Lysates. Five putative clones of prostromelysin were tested for stromelysin activity after activation with trypsin. Show here is the colorimetric assay data for stromelysin activity. Each box represents a time course of color development for one well in a 96-well plate. The substrate is Ac-Pro-Leu-Gly-[2-mercapto-4-methyl-pentanoyl]-Leu-Gly-OEt. Reactions were at room temperature for 30 minutes with readings every 11 seconds. Measurements were taken at 405 nm. The OD represented by the box height is 0.1.

Row A. Standard curve. Mature form of stromelysin (from Agouron)
Wells 1-6: 5 ug/ml, 2.5 ug/ml, 1.25 ug/ml, 0.625 ug/ml, 0.31 ug/ml, 0 ug/ml
Row B All the samples in this row were untreated.
Rows C & D. All the samples in these rows were treated with 25 uM trypsin for 10 min at 37C.
Row E & F. All the samples in these rows were treated with 4.16 uM trypsin for 10 min at 37C.
Row G & H. All the samples in these rows were treated with 1.0 uM trypsin for 10 min at 37C.
Column 1 (B thru H): 20 ug of crude lysate from clone strom1 (a pET 3d clone)
Column 2 (B thru H): 20 ug of crude lysate from clone strom3 (a pET 3d clone)
Column 1 (B thru H): 20 ug of crude lysate from vector alone (pET 28a)
Column 1 (B thru H): 20 ug of crude lysate from clone strom4 (a pET 28a clone)
Column 1 (B thru H): 20 ug of crude lysate from clone strom9 (a pET 28a clone)
Column 1 (B thru H): 20 ug of crude lysate from clone strom10 (a pET 28a clone)

This data shows that two of the clones, strom3 and strom9 can be activated with trypsin.

3d. Purification of his-tagged prostromelysin (our clone) on nickel columns.

We chose to work up the strom9 clone since it has a his-tag to facilitate purification while strom3 does not. Duplicate cultures of bacteria containing the strom9 expressor plasmid were grown up an induced as per 3c. Samples were collected every hour for 6 hours after induction. Lysates were prepared as described in 3c and analyzed by electrophoresis on polyacrylamide gels (Figure 13). On induction a

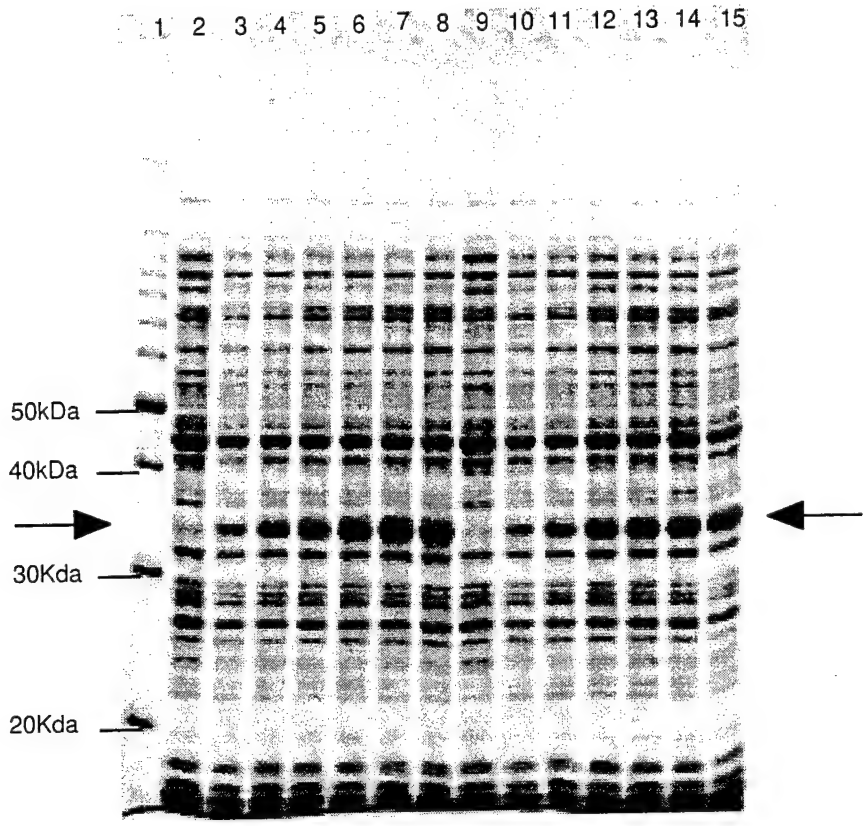
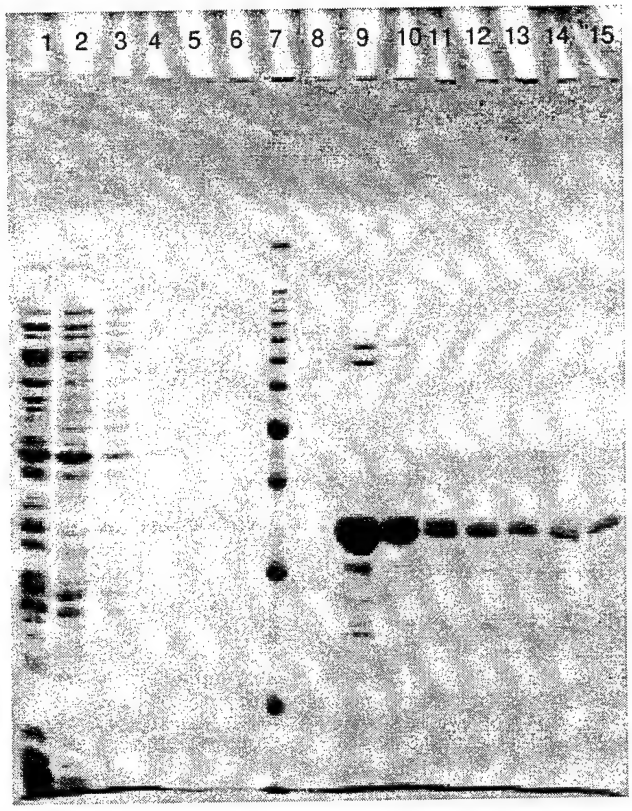


Figure 13. Induction of the Prostromelysin 9 clone. Duplicate cultures of the prostromelysin clone strom9 were induced and lysates produced as described in 3c. Samples taken every hour after induction were analyzed on a 12% polyacrylamide-1% SDS gel and stained with coomassie blue. The arrows indicate the new 33 kD band which is seen on induction.

33 kD band is seen (the predicted size is 32 kD). That is, these clones were constructed with C-terminal truncations to make the shortest possible 'pro' form of stromelysin, not the natural length prostromelysin.

To purify the soluble his-tagged prostromelysin from these cells a frozen pellet from 1250 ml of culture was thawed on ice and resuspended in 10 ml dHOH and the final solution was then made up to 0.2 mg/ml in lysozyme, 10 ug/ml DNase, 10 mM $MgCl_2$, and 1 mM PMSF. After 30 minutes on ice the solution was sonicated (5 30 second bursts with 30 second cooling between each). The solution was then cleared by centrifugation at 4C. The supernatant was then filtered through a 0.45 micron syringe filter and kept on ice while the column was prepared. 4 ml of his-bind resin (Novagen, Inc. Madison, WS) was put into a 50 ml centrifuge tube and spun in a tabletop centrifuge for 2 minutes at 1540 rpm. The 2 ml packed resin was then washed with three resin volumes of dHOH removing. The wash water by centrifugation. The column was then charged with nickel by washing with 5 resin volumes of 50 mM $NiSO_4$, removing the wash solution by centrifugation. Finally the column is washed with 3 resin volumes of 5 mM

imidazole, 500 mM NaCl, 20 mM Tris(pH 7.9). Lysate and resin were then mixed and placed on a shaker which gently agitated (50 rpm) the mix at 4C for 1 hour. The mixture was then centrifuged and the supernatant removed for analysis. The resin and bound prostromelysin was then washed five times for 10 minutes with 2.5 resin volumes of 60 mM imidazole, 500 mM NaCl, 20 mM Tris (pH 7.9) at 4C on the shaker. The final wash mixture was transferred to poly-prep chromatography column (BioRad, Hercules, CA) and the final wash collected as flow through. Prostromelysin was eluted using 14 mls of 1 M imidazole, 500 mM NaCl, 20 mM Tris (pH 7.9). 1 ml fractions were stored at 4C for analysis on 12% polyacrylamide-1% SDS gels (Figure 14).



1 non-binding flowthrough
2 Wash 1
3 Wash 2
4 Wash 3
5 Wash 4
6 Wash 5
7 10 kDa Marker
8 Elution fraction 1
9 Elution fraction 2
10 Elution fraction 3
11 Elution fraction 4
12 Elution fraction 5
13 Elution fraction 6
14 Elution fraction 7
15 Elution fraction 8

Figure 14. Purification of Prostromelysin (our clone strom9) on a Nickel Column. Non-binding, wash and elution fractions analyzed on a 12% polyacrylamide-1% SDS gel stained with coomassie blue. Arrow indicates size of his-tagged prostromelysin.

To determine the yield of protein from the column gel fractions were run on an electrophoresis gel on which was also run a series of known amounts of bovine serum albumin. Molecular Analyst software from BioRad (Hercules, CA) was used to create a standard curve using the BSA standards from which the

amount of prostromelysin in the various fractions could be determined (Figure 15). From this analysis we determined that we have recovered 2 BSA equivalent milligrams of prostromelysin per liter of culture.

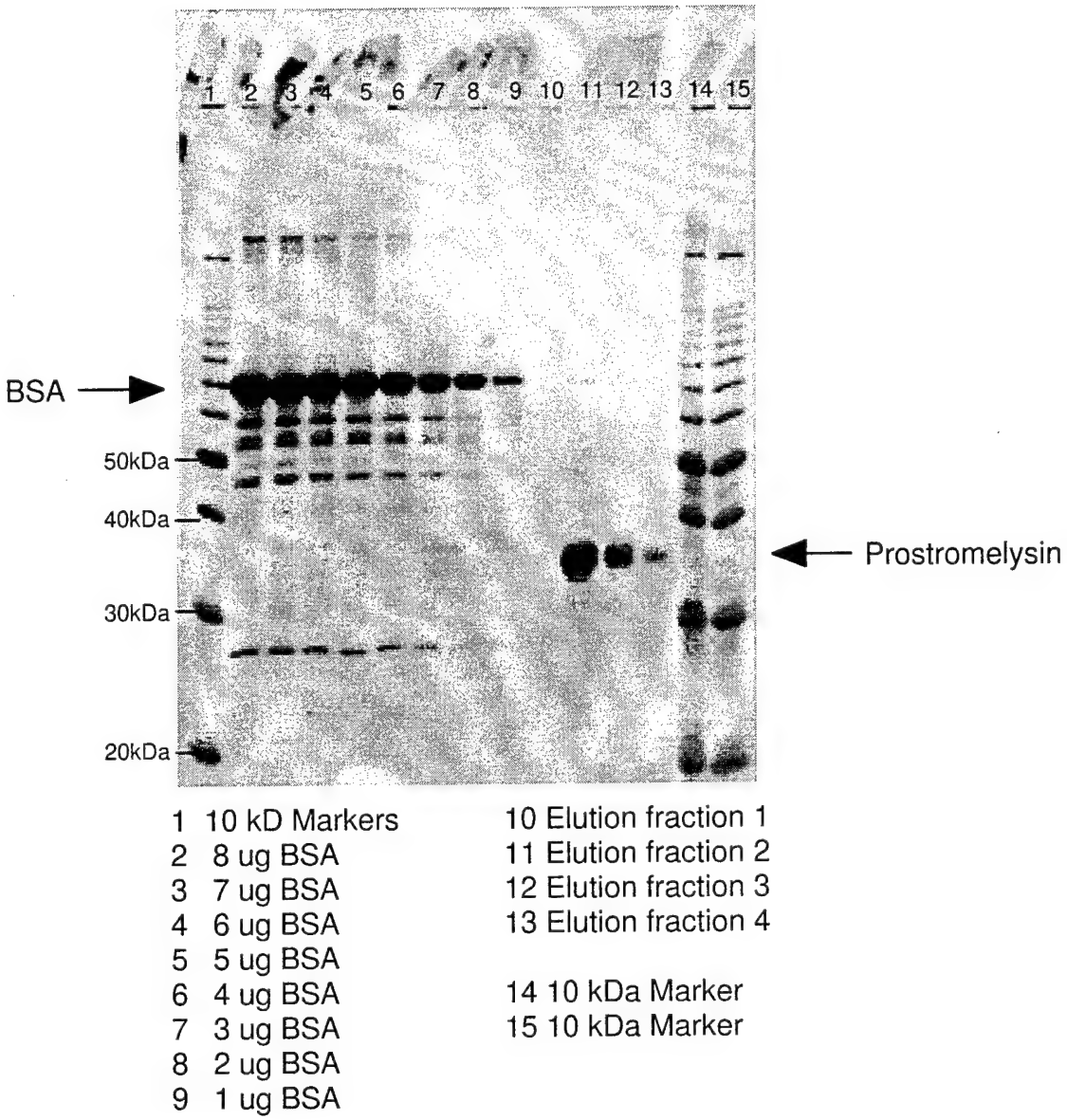


Figure 15. Quantitation of Protein Yield. Various known amounts of BSA were run on a 12% polyacrylamide-1% SDS gel to serve as a standards to quantitate the amount of protein recovered in the elution fractions from the nickel column. The Molecular Analyst software package was used to determine the amount of coomassie stain in the BSA fractions from which a standard curve could be created. Using the same software to determine the amount of coomassie in the prostromelysin bands in the elution fractions allowed us to determine the amount of BSA equivalent protein recovered.

4. Characterization of our Protein Engineering Framework

Eglin c is the framework we are using onto which we will graft the features necessary to create new inhibitors. In this project we are exploring the extent to which biophysical characterization is useful in hibitor construction. Both wt eglin c and his-tagged eglin c (the H_6 variant) are being used used as the

basis of libraries in our screening procedures. Therefore, we must understand both proteins before characterizing variants isolkated from these libraries. The equilibrium thermodynamics of denaturation is most often interpreted with a reversible two-state model,

$$N \longleftrightarrow D$$

where N is the native state and D is the thermally denatured state. The reaction's equilibrium constant, K_D, is related to the thermodynamic parameters via the Gibbs-equation:

$$-RT\ln K_D = \Delta G_D = \Delta H_D - T\Delta S_D,$$

where $\Delta G_{D,T}$ is the free energy of denaturation, defined s the stability, $\Delta H_{D,T}$ is the enthalpy of denaturation, $\Delta S_{D,T}$ is the entropy of denaturation and T is the absolute temperature. Because there is a difference in heat capacity between the N and D, $\Delta H_{D,T}$, $\Delta S_{D,T}$, (and hence $\Delta G_{D,T}$) are temperature dependent. This temperature dependence is described by the equations

$$\Delta H_{D,T} = \Delta H_m + \Delta C_P(T - T_m),$$

$$\Delta S_{D,T} = \Delta S_m + \Delta C_P \ln\left(\frac{T}{T_m}\right)$$

$$y_{N,T} = m_N T + b_N$$

where T is the temperature of interest and the subscript *m* refers to a reference temperature. The terms m_N and m_D are the changes in CD with respect to temperature of N and D, respectively. Likewise, b_N and b_D are the CD of N and D extrapolated to 0 K.

We define the reference temperature where K_D is unity and, therefore, ΔG_D. is zero. Combining the above two equations, defining ΔS_m as $\Delta H_m/T_m$ and simplifying gives integrated for of the Gibbs-Helmholtz equation.

$$\Delta G_{D,T} = \Delta H_m\left(1 - \frac{T}{T_m}\right) - \Delta C_P[(T_m - T) + T\ln\left(\frac{T}{T_m}\right)]$$

Inspection provides the cardinal thermodynamic parameters, ΔH_m, T_m, and ΔC_p, that we must obtain to determine eglin c stability at any temperature. We have obtained these parameters by examining the far-UV circular dichroism (CD) as a function of temperature and by using differential scanning calorimetry (DSC).

Using CD to obtain the cardinal parameters CD in the far ultraviolet (~220 nm) gives information about secondary structure. The simplest way to monitor denaturation is to follow the CD as a function of temperature. By defining the low temperature baseline as representing fully native protein and defining the high temperature baseline as representing fully denatured protein, we obtain a fraction denatured-versus-temperature plot. Assuming the denaturation reaction is two-state, the fraction of denatured protein at temperature T, $\alpha_{D,T}$, is

$$\alpha_{D,T} = \frac{A_{\lambda,T} - y_{N,T}}{y_{D,T} - y_{N,T}}$$

where the portions of the signal, $A_{\lambda,T}$, due to the N and D baselines (y_N and y_D, respectively) are linear functions of temperature.

The equilibrium constant for denaturation, $K_{D,T}$, is

$$K_{D,T} = \frac{[D]}{[N]} = \frac{\alpha_{D,T}}{1 - \alpha_{D,T}}.$$

T_m is defined above and ΔH_m is obtained using the vant Hoff equation

$$\Delta H_m = \frac{\delta(-R\ln K_{D,T})}{\delta(1/T)}$$

We can then determine the heat capacity change for denaturation, ΔC_p, using Kirchoff's law,

$$\Delta C_p = \frac{\delta \Delta H}{\delta T}.$$

In theory ΔC_p can be obtained from the curvature of vant Hoff plots. In practice, however, accurate $K_{D,T}$ values are available over a narrow temperature range, making curvature difficult to detect (Cohen & Pielak, 1994). Instead, Pfeil and Pivalov (1976) assume that ΔH_D is pH independent. This assumption is reasonable because the ionization heat of carboxyl groups, the protein groups which ionize in the pH region we examined, is ~0 (Edsall & Wyman, 1958). Given this assumption, ΔC_p can be determined by obtaining ΔH_m and T_m at different pH values because the slope of such a plot has the form of Kirchoff's law. Finally, we can obtain the number of groups that ionize on denaturation ($\delta\nu$) by fitting T_m-versus-pH data

$$\Delta \nu = \frac{\Delta H_m}{2.3RT^2} \frac{\delta T_m}{\delta pH}$$

(Privalov &Ptitsyn, 1969; Ptitsyn & Birshtein, 1969).

Using DSC to obtain the cardinal parameters DSC measures the constant-temperature heat capacity, C_p, as a function of temperature and the resulting plots are called thermograms. ΔH_{cal} is the area under thermogram from the defining equation

$$\Delta H = \int C_p dT.$$

It is important to note the fundamental difference between calorimetric and vant Hoff enthalpies: the former require knowledge of the protein concentration the latter requires do not. If denaturation is a two-state process, the temperature at which C_p is a maximum is T_m and a H_T-versus -T plot has the same shape as a α-versus-T plot. Therefore, the vant Hoff enthalpy can be obtained from DSC data in the same manner as it is obtained from CD data.
In theory, ΔC_p can be obtained directly from the difference in ΔC_p between the pre- and post-transition baselines. This regimen, however, is not dependable in our hands and we determined ΔC_p from DSC data as a function of pH by obtaining ΔH_{cal} and ΔH_{vH} (from the calorimetry) at several pH values as described above.

4a. Determination of basic thermodyamic properties of eglin c and H_6 eglin.

Eglin c and the H_6 variant were expressed using *Escherichia coli* strain BL21(DE3)pLysS and the pET17b (Novagen) vector. This vector utilizes a T7 promoter and is maintained with carbenicillin. With this system, ~40% of the expressed protein is eglin c. The purification exploits one of eglin c's unusual properties, resistance to thermal denaturation. Cell extracts were cleared by boiling, followed by size-exclusion chromatography on Sephadex G-75. Our isolation/purification method yields 100-200 mg of pure eglin c per liter of saturated culture.

Thermodynamic data were acquired in 50 mM glycine-HCl. Except for concentration-dependence experiments, CD and DSC data were acquired at protein concentrations of 60 and 90 μM, respectively. CD data were acquired in 1.00 mm quartz cells at 237 nm, a wavelength that gives the maximum difference in signal upon denaturation, using an Aviv Model 62DS spectropolarimeter equipped with a thermostated, five-position sample changer. The effective scan rate when all five positions were used was 0.1 K min^{-1}. CD data were processed as described by Cohen and Pielak (1994). Calorimetry data were acquired at a 60 K min^{-1} scan rate using a Microcal MCS2 DSC and processed using Origin software.

Reversibility First, we had to find conditions where repeated denaturation of the same sample gives overlaying α-versus-T plots. In addition, the conditions must allow complete observation of N-D transitions between the freezing point and the boiling point of our buffer. For both proteins the criteria are met between pH 1.5 and 4 for CD data. Several representative data sets are shown in Figure 16. Above pH 4 the wt protein precipitates, probably because we approach its isoelectric point. It will be interesting to determine if the H_6 variant, which has a higher isoelectric point, also precipitates.

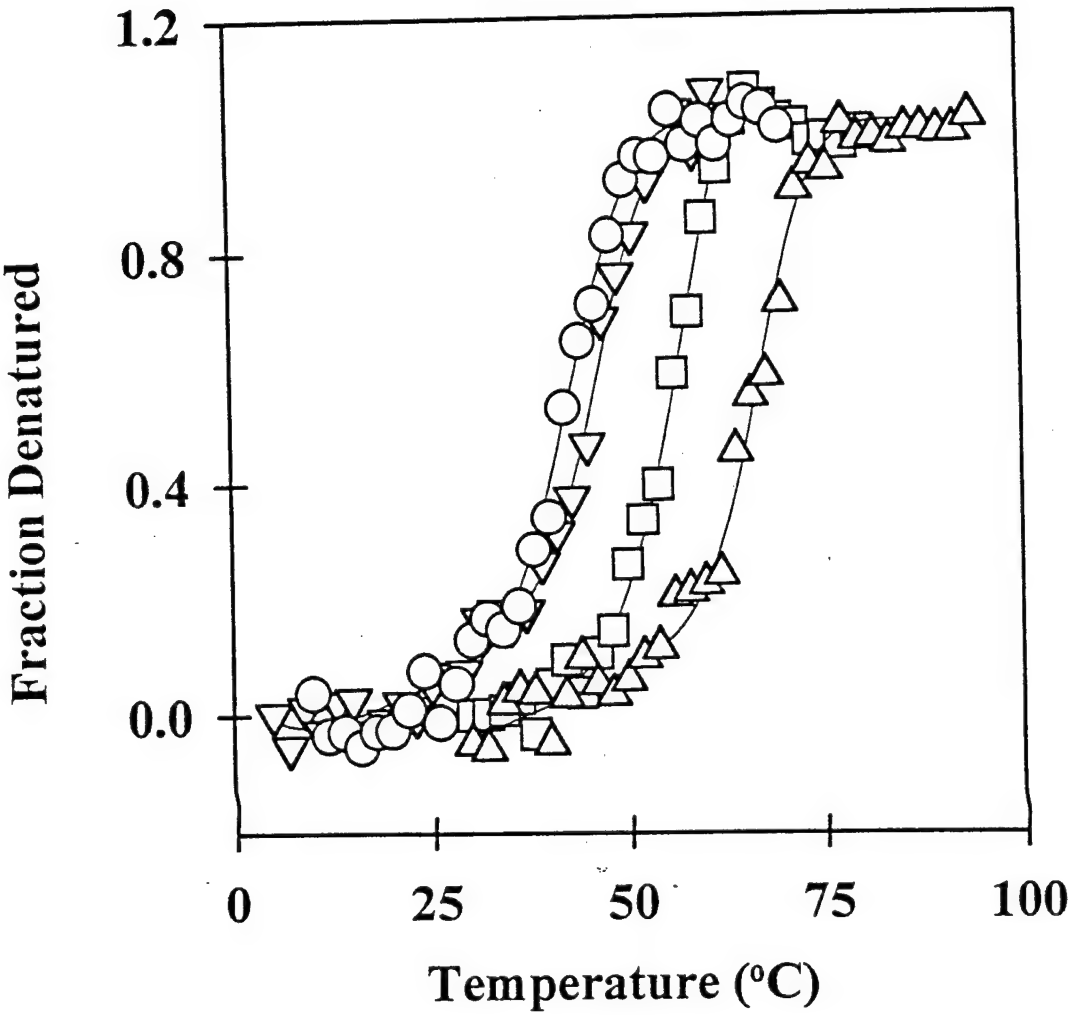


Figure 1. Fraction denatured-versus-temperature. Denaturation plots for H_6 eglin c using circular dichroism in 50 mM glycine buffer at pH 1.6 (O), pH 2.0 (∇), pH 2.6 (□), and 3.4 (Δ).

analysis	ΔH_D kcal mol^{-1}	T_m (pH 3.0) K	ΔC_P kcal mol^{-1} K^{-1}	$\Delta G_{D,300K,pH3}$ kcal mol^{-1}
cal	21.1 ± 12.6	na*	1.0 ± 0.2	4.4 ± 1.2
vH	41.0 ± 3.9	338.0 ± 0.3	0.7 ± 0.1	6.1 ± 0.3
cal	26.9	na	0.84	4.5
vH	32.6	336.3	0.73	5.1

***na, not applicable.**

Table 4. Cardinal thermodynamic parameters for wt eglin c from DSC data. Uncertainties are standard deviations of the mean from replicate experiments.

We also tested reversibility in the DSC. Repeated temperature scans give identical T_m and ΔH_{vH} values, but ΔH_{cal} decreases with repetition. We interpret this observation to mean that protein is lost in each scan, but the protein that remains in solution denatures reversibly. In summary, there are two kinds of reversibility, thermodynamic and structural. Thermodynamic reversibility refers to the fact that the ΔH_{vH} from CD and DSC are the same for repeated scans. Structural reversibility refers to the fact that ΔH_{cal} decreases with repeated scans. We are examining regimens for extrapolating the calorimetric data back to zero time.

Two-state behavior Our analysis **assumes** that thermal denaturation is a two-state process. We tested this assumption as follows. We determined ΔH_D both indirectly using vant Hoff analysis of CD and DSC data and directly by DSC. Using calorimetry we need not rely on assumption about the reaction to obtain ΔH_D. Therefore, if thermal denaturation is a two state process, $\Delta H_{D,vH} = \Delta H_{D,cal}$

As shown in Table 4, $\Delta H_{D,vH}/\Delta H_{D,cal}$ is >1. (We are collecting data to determine the parameters from CD data and we are working to lower the uncertainty in ΔH_{cal}..) There are at least three explanations for this observation: intermolecular cooperation, loss of protein in the DSC, and part of the protein does not unfold on denaturation. Intermolecular cooperation involves aggregation of either the native or the denatured state (Sturtevant, 1987). However, the cardinal parameters from neither CD nor DSC data are concentration dependent; an observation also made by Bae and Sturtevant (1995). Another possibility is that ΔH_{cal} is too small because of protein loss. In fact, we discard the first thermogram of a run because it gives unreliable parameters. We need to find the source of this behavior. If the source is loss of protein the extrapolation experiment described above will give the answer. Third, a portion of the protein may not unfold on denaturation; either a part of the protein is unfolded in both states or a portion fails to unfold. In both cases the ΔH_{vH} will be greater than ΔH_{cal} because the former gives information only about the portion of the protein that unfolds, but the latter gives information about the enthalpy change per mole of protein. Put another way, we divide the heat of denaturation by the number of moles of protein to ΔH_{cal}, but the protein concentration is not required to obtain ΔH_{vH} . If a part of the protein is unfolded in both states, we expect $\Delta H_{vH}/\Delta H_{cal}$ to increase upon addition of the his-tag. However, the opposite is observed (Tables 4 and 5). On the other hand, if a part of the protein fails to unfold upon denaturation we might expect to see residual structure in the high-temperature far-UV CD spectrum. In fact we observe such structure; the high temperature spectrum shows β-sheet-like features, not the expected random coil features. We are going to compare opening free energies from amide proton exchange data to denaturation free energies from CD and DSC data. We have shown that such structure exists in ferricytochrome c (Marmorino et al., 1993; Betz et al., 1996).

The cardinal thermodynamic parameters Parameter sets for both proteins from CD and DSC data are shown in Tables 4 and 5. As discussed above, the vant Hoff enthalpy is larger than the calorimetric enthalpy. We are working to explain this difference. We can also compare our data on wt eglin c to those obtained by Bae and Sturtevant (1995). There is excellent agreement between the CD and DSC data.

analysis	ΔH_D kcal mol^{-1}	T_m (pH3.0) (K)	ΔC_P kcal mol^{-1} K^{-1}	$\Delta G_{D,300K,pH3}$ kcal mol^{-1}
DSC, cal	30.7 ± 12.6	na	0.9 ± 0.2	5.4 ± 1.2
DSC, vH	38.8 ± 3.9	338.4 ± 0.3	0.8 ± 0.1	6.1 ± 0.3
CD, vH	20.7 ± 7.5	338.5 ± 1.6	1.1 ± 0.1	4.7 ± 1.7

***na, not applicable.**

Table 5. Cardinal thermodynamic parameters for H6 eglin c from CD and DSC data. Uncertainties are standard deviations of the mean form replicate experiments.

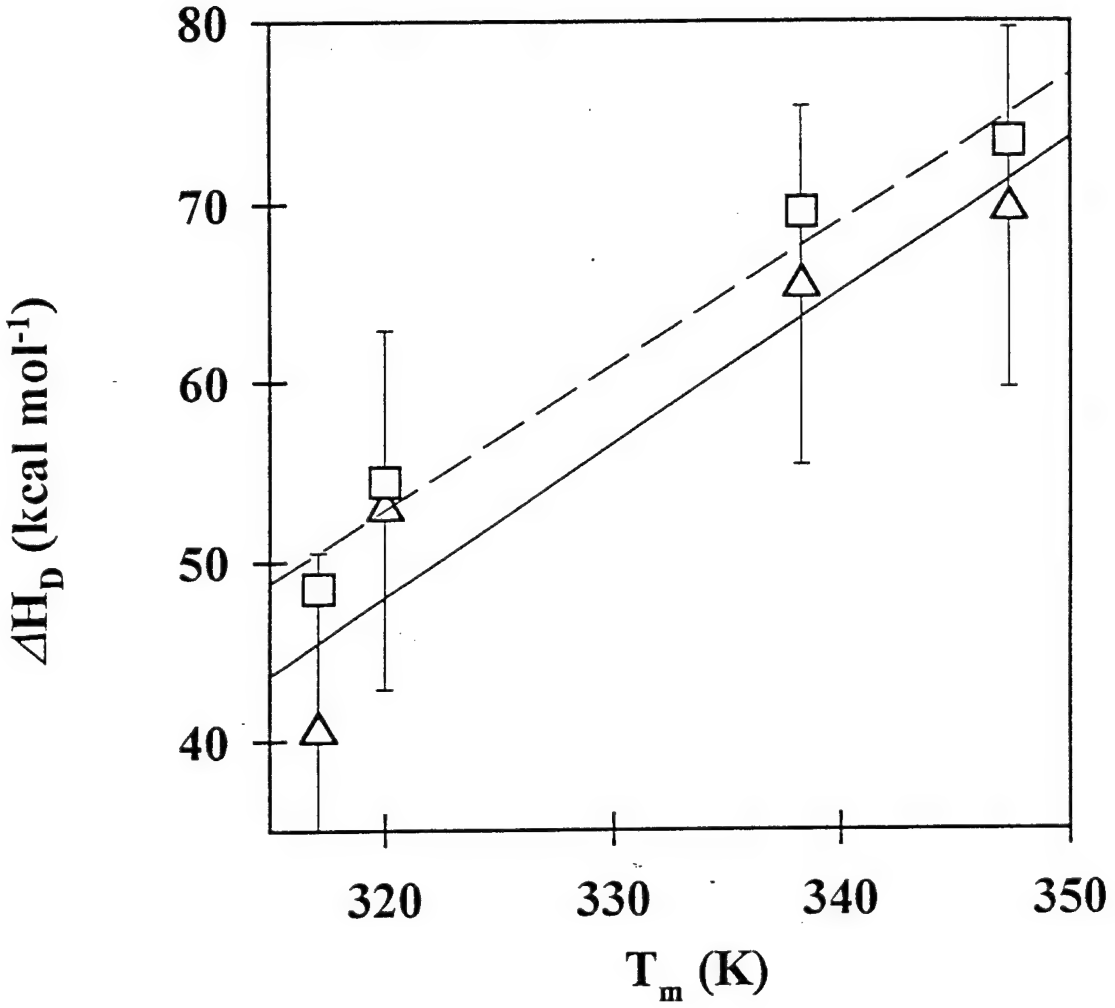


Figure 17. ΔH_D versus T_m for H_6 eglin c from (❑), van't Hoff and (Δ), calorimetric analysis of DSC data. The uncertainties in ΔH_{vH} and ΔH_{cal} are the standard deviation of the mean obtained from four repitions of the same experiment. The uncertainty in ΔH_{vH} is smaller than the symbol and uncertainty in ΔCp is from non-weighted linear regression.

Figure 17 shows ΔH_D-versus-T_m plots for the H_6 variant. The plots are well described by a straight lines. The resulting ΔC_P values are listed in Table 2. Similar plots were made for the wt protein (data not shown) and resulting ΔC_P values are given in Table 1 along with the value obtained by Bae and Sturtevant (1995). The values agree.

Effect of the his-tag. We next examined the data for the H_6 variant to determine if this protein is useful for library construction. That data shows that the addition of the his-tag has only a small effect on T_m and ΔH_D, and no effect on ΔC_P. Therefore, the addition does not effect stability, ΔG_D at pH 3.

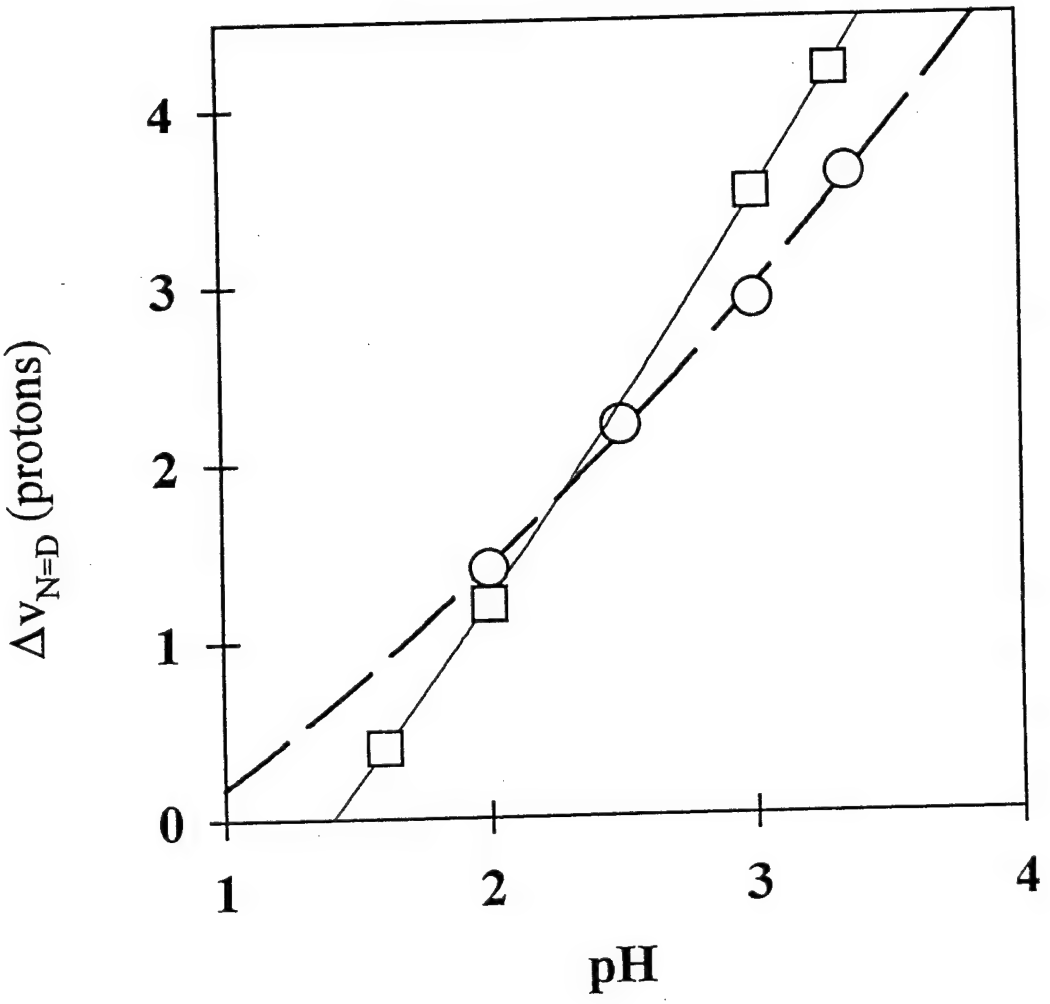


Figure 18. Δν-versus-pH plot for wt (○) and H_6 eglin c (□) at 300K in 50 mM glycine buffer. The uncertainty in Δν was calculated by propagation of error analysis of $\Delta\nu = \Delta H_m(\delta T_m)/(2.3RT_m^2)(\delta pH)$.The uncertainty is smaller than the symbol size. The curves are of no theoretical significance.

Figure 18 shows the change in the number of protons taken upon denaturation, Δν. Both proteins gain between 1 and 4 protons between pH 1.5 and 3.4. This behavior is typical of most proteins. The addition of the his tag decreases Δν slightly. Finally, we combined all the data to make a plot ΔG_D at 300 K versus pH (Figure 19).

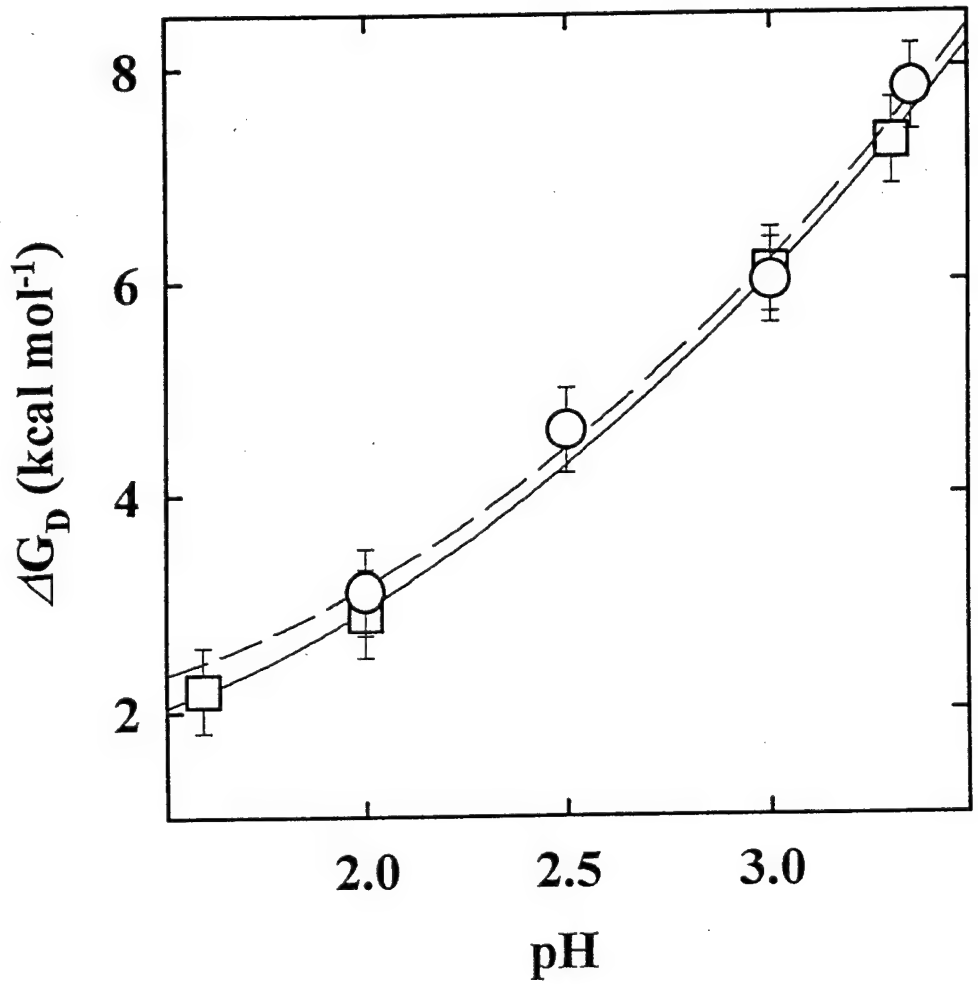


Figure 19. ΔG_D-versus-pH for wt (O) and H_6 eglin c (❑) at 300K in 50 mM glycine buffer. The uncertainty in ΔG_D is from propagation of error analysis on the Gibbs-Helmholtz equation (Cohen & Pielak, 1994). The curves are of no theoretical significance.

The wt protein is more stable than the H_6 variant at all pH values, but the effect increases with increasing. However, the maximum decrease is stability is less than 1 kcal mol^{-1}.

Summary Although we must complete these experiments to refine out data, it is clear that eglin c and its H6 variant provide useful thermodynamic data; even though the his-tag has a small destabilizing effect, the H_6 variant can be used as the basis of random mutant libraries.

4b. Exploration of the eglin c structural determinants.

Anfinsen and coworkers (1973) showed that the information necessary for a protein to fold into its native, active form is contained in its amino acid sequence. Given this knowledge, it should be possible to predict the sequences that dictate a specified fold. The relationships between protein primary structure and higher-order structure, stability and function, however, remain an enigma. As part of our efforts to understand the eglin c structure and use it for a framework for inhibitor construction we are exploring a strategy to evaluate proposed hyptheses concerning protein structure. We will test this strategy by examining an evolutionary conserved amino acid pattern for its ability to direct α-helix formation in eglin c.

Using FASTA (Pearson, 1980), 20 proteins were found that possess >30% sequence identity with eglin c. These sequences were aligned using the program MulAlignment (Computational Biochemistry Research Group, Zurich, Switzerland). The alignment of the α-helical segment in 20 eglin c homologs follows a binary pattern, '**onnoo**' (Figure 20). A non-random pattern emerges when residues are divided into two classes: nonpolar (**n**), Leu, Ile, Val, Phe and Met (LIVFM) and polar(**o**) non-LIVFM. Vasquez et al. (1993) found that they could encode structures using this same binary amino acid pattern among natural proteins. So this is the pattern that we will test. We used the crystal and solution structures of eglin c (Bode et al., 1986; McPhalen & James, 1988; Hyberts & Wagner, 1990) to calculate each residue's solvent exposed surface area (SASA) and hence the polar context of the wt residues.. Surprisingly the α-helix residues in wild-type (wt) eglin c contain two mismatches with the binary encoding pattern. We expect that eglin c variants that do conform to the pattern will be even more stable than wt.

A library will be made utilizing the evolutionary-dictated pattern. Variants will be expressed and screened for activity. Active variants will be studied using circular dichroism spectropolarimetry (CD) and differential scanning calorimetry (DSC). CD spectra will provide secondary structure data. Both CD and DSC will provide equilibrium thermodynamic data. Free energies of denaturation, ΔG_D, will be compared to that of wt eglin c to ascertain the ability of the pattern to dictate a stable fold.

In summary, the **main objective** is to determine if a binary pattern from the evolutionary record dictates formation of the eglin c α-helix. To test this hypothesis, we will:

A. **Produce and screen a library** using the α-helix pattern defined by eglin c homologs. These variants will be expressed and screened using a high through-put subtilisin inhibition assay.
B. **Characterize secondary structure and stabilities of wt and variants** using far-UV CD and DSC.

Results.

PCR mutagenesis was used to insert a Bam HI restriction site at the N-terminal end of the helix cassette in pET17b (Kuipers et al.,1991; Novagen, Madison, WI) (Figure 21). The Aat II restriction site at position 3497 was then removed by digesting the plasmid with Aat II for 15 min and then heat inactivating the enzyme. The reaction mixture was then treated with the Klenow fragment of *E. coli* DNA polymerase for 2 h followed by overnight blunt-end ligation. The DNA was transformed and mutant colonies were isolated and identified by DNA sequene analysis to yield pET17bBA (Figure 22). We are inserting an Hpa I site and two stop codons between the Bam HI and Aat II sites as a cassette to produce the "cassette ready plasmid", pJCW1 (Figure 23).

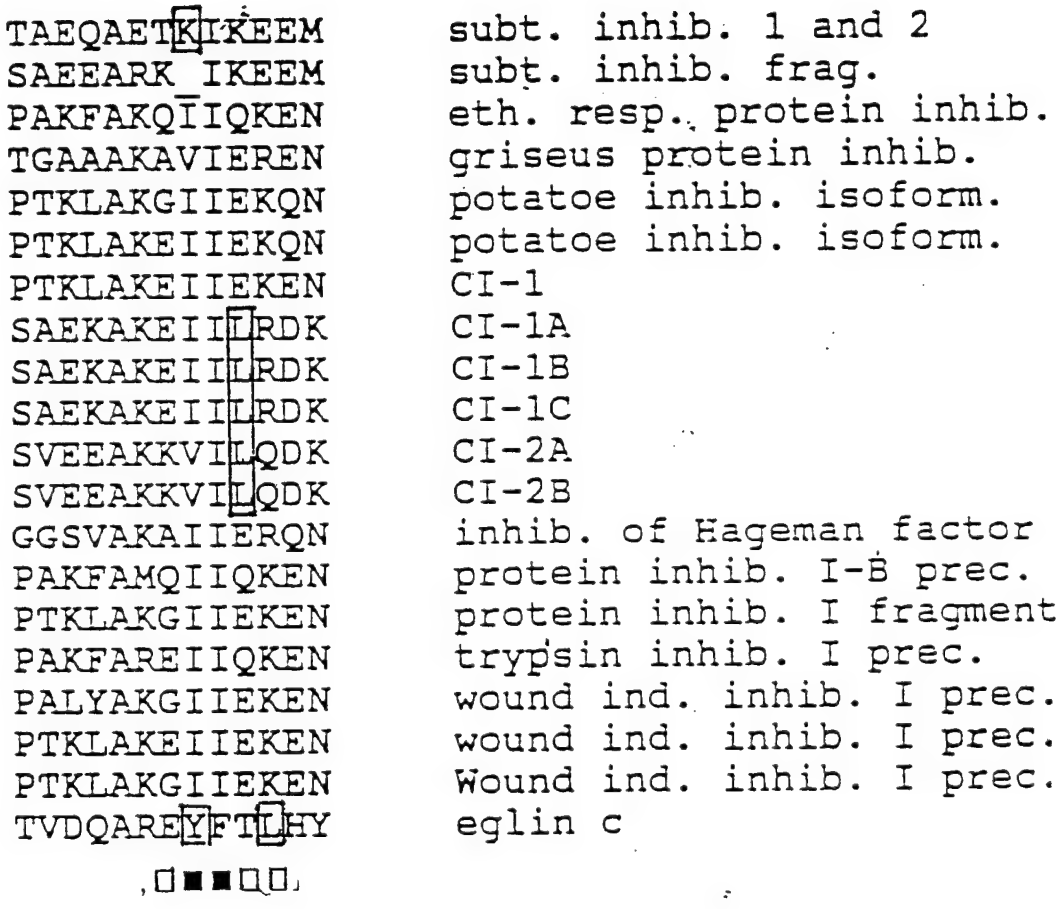


Figure 20. Eglin c homologs. (❑) polar and (■) nonpolar residues conserved. The residues defined as non-polar are LIVFM. The polar residues are defined as non-LIVFM. Nonconforming residues are boxed.

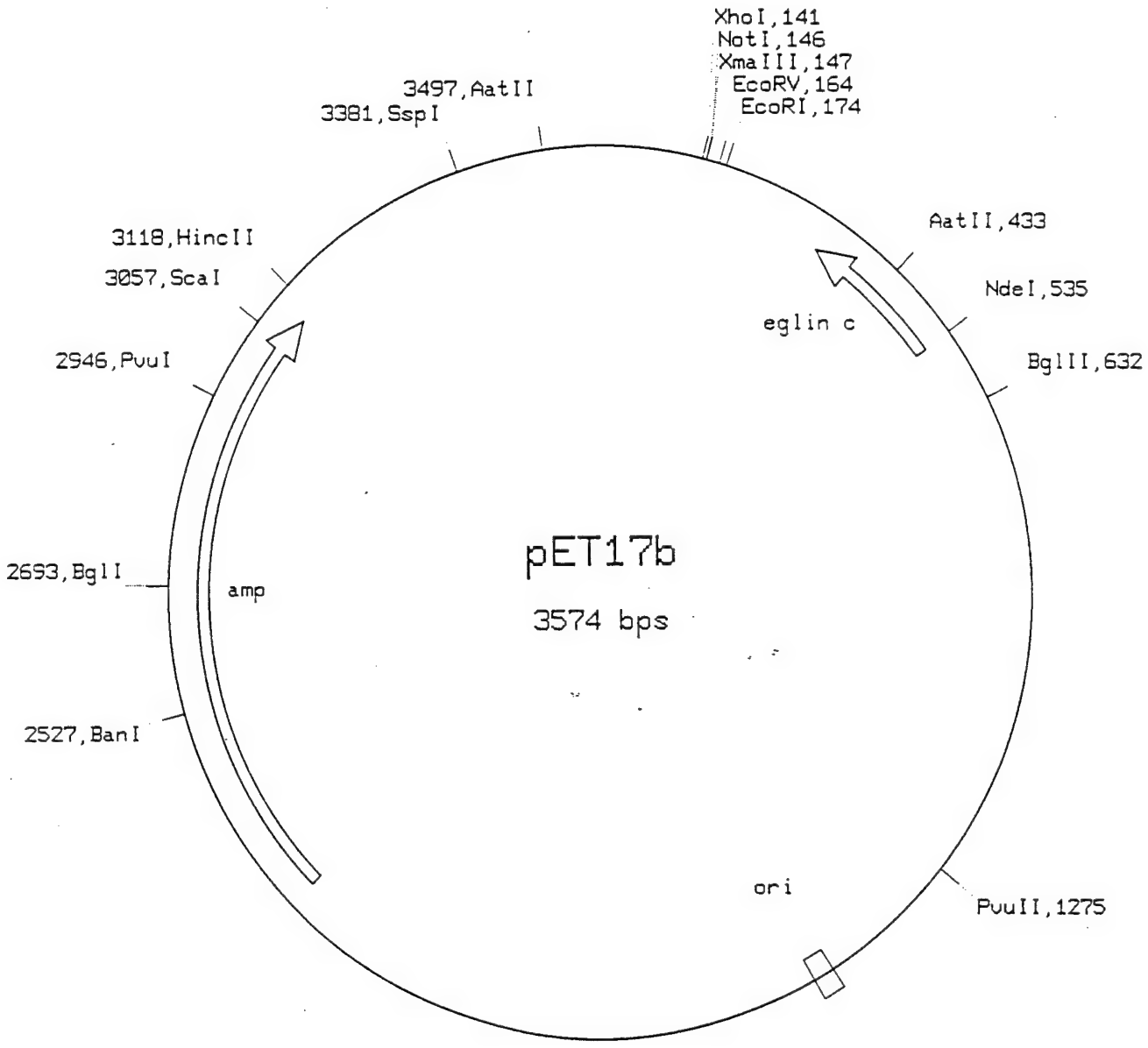


Figure 21. Map of pET17b. This is the starting plasmid for our construction of another plasmid (next figure) more suitable for library construction involving the eglin α-helix. PCR mutagenesis was used to insert a Bam HI restriction site at the N-terminal end of the eglin α-helix cassette in pET17b just clockwise from the Aat II-433 site. The Aat II restriction site at position 3497 was then removed by partially digesting the plasmid with Aat II for 15 min and then treating with the Klenow fragment of *E. coli* DNA polymerase followed by overnight blunt-end ligation. The DNA was transformed and mutant colonies were isolated and clones missing the Aat II-3497 site identified by DNA sequence analysis to yield pET17bBA (Figure 22).

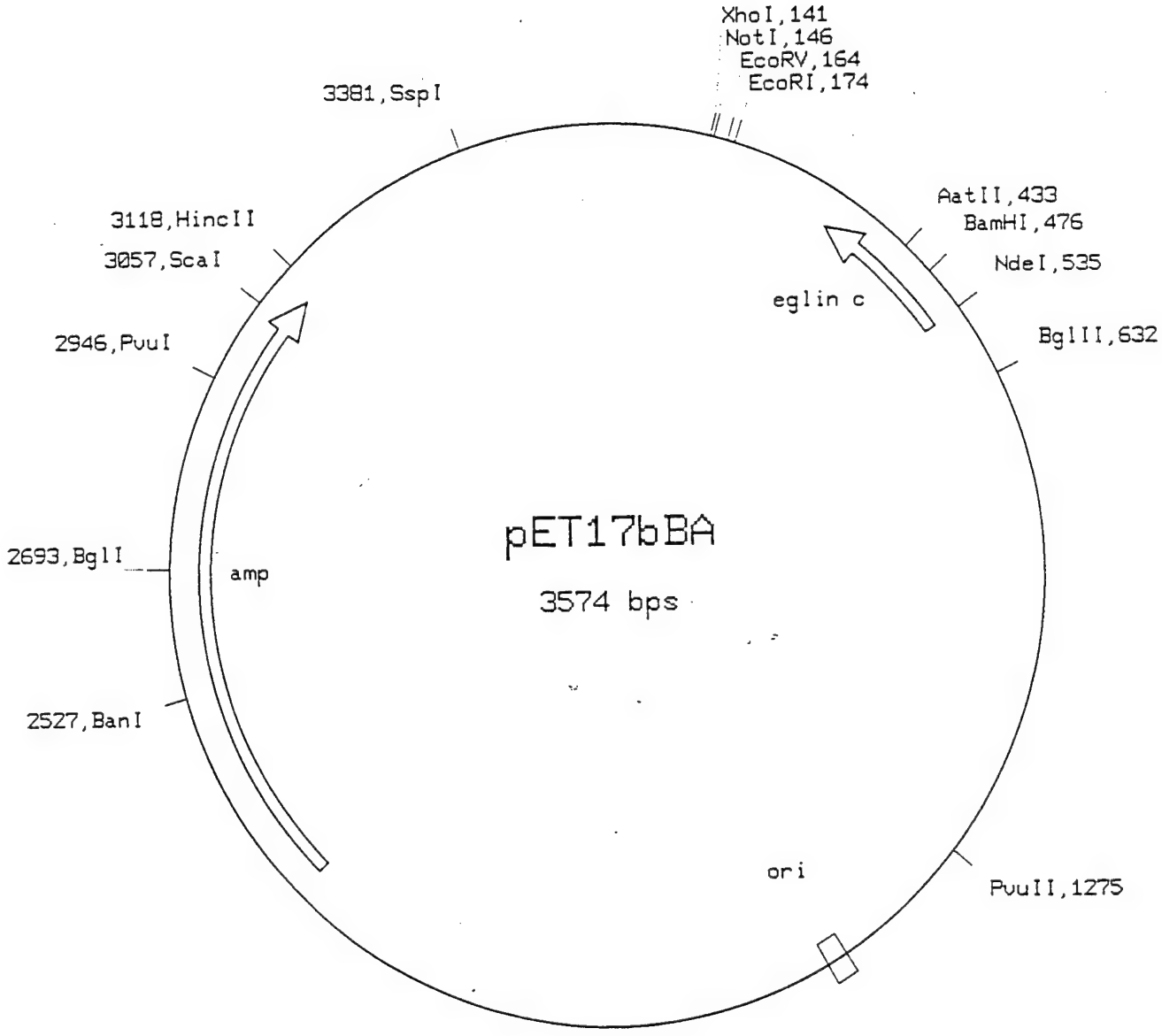


Figure 22. Map of pET17bBA. This is the resultant plasmid with restriction sites flanking the eglin c a-helix. It was constructed from pET17b as per the text or figure 21 legend.

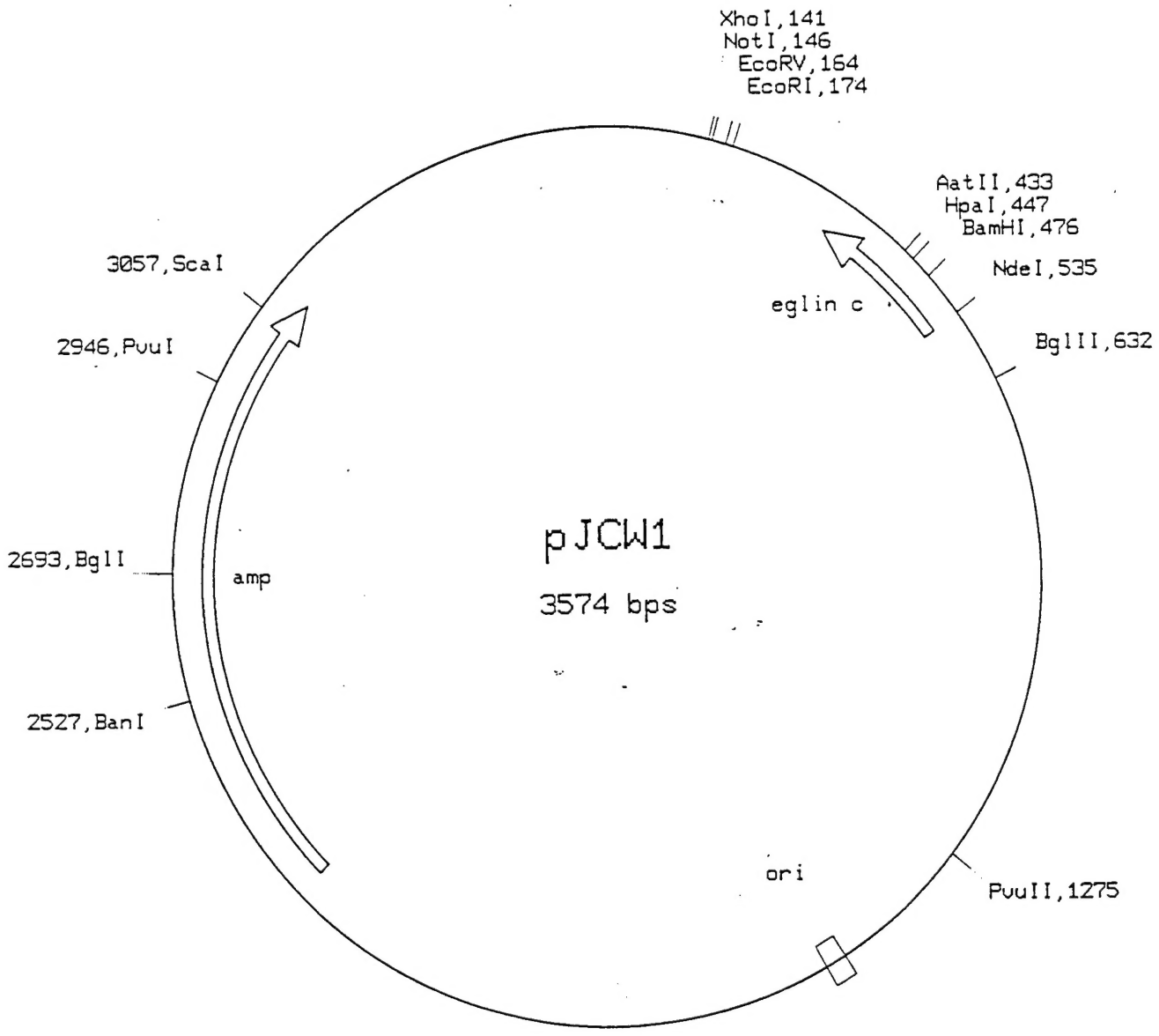


Figure 23 Map of pJCW1. This is the plasmid which will be used for library construction. Library construction is often plagued by a large fraction of clones which represent the unmodified wild type plasmid. To render the starting vector inactive in terms of eglin protein production we are inserting an Hpa I site and two stop codons between the Bam HI and Aat II sites. The two stop codons will insure that no **protein of** the elgin size is produced and the Hpa I site will be used to digest the library construction since only the original vector should be digested.

Based on this analysis a helix cassette based on the evolutionary conserved binary pattern, Q O O Q Q, has been designed where O = Leu, Ile, Val, Phe, and Met and Q = Asp, Glu, Lys, Asn, Gln, His, Arg, and Ser. Three putative stabilizing interactions at the helix N-terminus that will be preserved in the library. First, Thr17 and Gln20 are involved in reciprocal H-bonds, called N-capping (Harper & Rose, 1993; Bode et al., 1986). Second, the side chain of Asp19 accepts a H-bond from its amide proton, stabilizing the helix dipole (Bode et al., 1986). Third, Ala21 is invariant (Figure 20). Comparison of the SASAs of eglin c's amino to the natural pattern, suggests that the designed binary pattern will direct helix formation. The helix side chains, Asp, Arg, Glu, Leu and His, are 87%, 56%, 68%, 69%, and 51% solvent exposed, respectively. This degree of exposure is beneficial for polar side chains because they are stabilized by interaction with water. Therefore, changing these residues to Asp, Glu, Lys, Asn, Gln, and His should give a stable, active protein. The Tyr and Phe side chains are 16% and 0.1% exposed. The burial of nonpolar residues is stabilizing because the protein interior provides a nonpolar environment and, therefore, changing these residues to Leu, Ile, Val, Phe, and Met should provide a stable and active protein. We are in the process of constructing this library.

CONCLUSIONS

The project as originally defined, that is, to work out a methodology to build protein-based inhibitors against proteins involved in metastasis, remains on track. Most importantly we have shown that the basic approach is sound by finding a variant of one of the framework molecules (teglin) which binds to a heterologous target, papain. We are using papain as a model target until sufficient stromelysin is produced to serve as a target.

To work around the fact that our originally chosen framework molecule did not work in the phage display system, we have constructed a circularly permuted form of eglin c, which we are calling peglin, and shown that this does bind to its appropriate target, subtilisin.

We have identified a clone of his-tagged pro-stromelysin which produces a molecule from which we can extract active stromelysin. Our first experiments suggest that the yield from liter cultures is sufficient to produce enough material to be used in the phage display system.

We have pursued experiments to collect baseline thermodynamic data concerning the framework molecules to be used in this project.

We have carried out a structural analysis of eglin c which will allow us to test a strategy for assessing structural rules. These rules should make it more easy for us to design our inhibitor libraries.

REFERENCES

Anfinsen, C B . Principles that govern the folding of protein chains. *Science* **181,** 223-230 (1973).

Bae, S-J, Sturtevant, J.M. Thermodynamics of the thermal unfolding of eglin c in the presence and absence of guanidinium chloride. *Biophys. Chem.* **55**, 247-252 (1995).

Betz, S.F., Marmorino, J.L., Saunders, A.J., Doyle, DF, Young, GB, Pielak, GJ . *Biochemistry* **35**, 7422-7428 (1996).

Bode, W, Papamokos, E, Musil, D, Seemhler, U, Fritz, H. . Refined 1.2 D crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. *EMBO J.* **5**, 813-818 (1986).

Cohen, D.S., Pielak, G.J. . Stability of yeast iso-1-cytochrome *c* as a function of pH and temperature *Protein Sci.,* **3**, 1253-1260 (1995).

Edsall, J.T. Wyman, J. . *Biophysical Chemistry,* vol. **1**, New York, Academic Press, pp. 452-453 (1958).

Harper; ET, Rose, GD . Helix stop signals in proteins and peptides: the capping box. *Biochemistry. 32*, 7605-7609 (1993).

Hyberts, SG, Wagner, G. . Sequence-Specific ^{1}H NMR Assignments and secondary structure of eglin c. *Biochemistry* **29,** 1465-1474 (1990).

Kuipers, OP, Boot, HJ, de Vos, W M . Improved site-directed mutagenesis methods using PCR. *Nucleic Acids Research* **19.** 4558 (1991).

Leatherbarrow, R. J. and H. J. Salacinski. Design of a small peptide-based inhibitor by modeling the active-site region of barley chymotrypsin inhibitor 2. *Biochemistry* **30** 10717-10721. (1991)

Marcy, A.I., Eiberger, L.L., Harrison, R., Chan, H.K., Hutchison, N.I., Hagmann, W.K., Cameron, P.M., Boulton, D.A., and Hermes, J.D. Himan fibroblast stromelysin catalytic domain: expression, purification, and characterization of a C-terminally truncated form. *Biochemsitry* **30** 6476-6483 (1991)

Marmorino, JL, Auld, D.S., Betz, SF, Doyle, DF, Young, GB, Pielak, GH . Amide proton exchange rates of oxidized and reduced *Saccharomyces cerevisiae* iso-1-cytochrome *c*. *Protein Sci.* **2,** 1966-1974 (1993).

McPhalen, C A, James, M. N. . Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin carlsberg and CI-2-subtilisin novo. B*iochemistry* **27,** 6582-6598 (1988).

Muller, D., Quentin, B., Gesnel, M., Millon-Collard, R., Abecassis, J., Breathnach, R. The collagenase gene family in humans consist of at least four members. *Biochem. J.* **253** 187-192 (1988)

Pearson, W R, . Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183**, 63-99 (1990).

Pfeil, W, Privalov, PL . Thermodynamic investigations of proteins III. Thermodynamic description of lysozyme. *Biophys. Chem.* **4,** 41-50 (1976).

Privalov, PL, Ptitsyn, OB . Determination of stability of the DNA double helix in an aqueous medium. *Biopolymers* **8,** 559-571 (1969).

Ptitsyn, OB, Birshtein, TM . Method of determining the relative stability of different conformational states of biological macromolecules. *Biopolymers* **7,** 435-445 (1969).

Seemhler, U, Meier, M, Ohlsson, K, Muller, H., Fritz, H. . Isolation and characterization of a low molecular weight inhibitor (of chymotrypsin and human granulocytic elastase and cathepsin G) from leeches. *Hoppe-Seyler's Z. Physiol. Chem.* **358,** 1105-1117 (1977).

Sturtevant, JM Biochemical applications of differential scanning calorimetry. *Annu. Rev. Phys. Chem.* **38**, *46-488* (1987).

Vasquez, S, Thomas, C, Lew, R A., Humphreys, R. E. . Favored and suppressed patterns of hydrophobic and nonhydrophobic amino acids in protein sequences. *Proc. Natl. Acad. Science U.S.A.* **90,** 9100-9104 (1993).

Weingarten, H. and Feder, J. Spectrophotometric assay for vertebrate collagenase. *Anal. Biochem.* **147** 437-440 (1985)

Ye, Qi-Zhunag, Jonhson, L.L., Hupe, D.J., and Baragi, V. Purification and characterization of the human stromelysin catalytic domain expressed in *Esherichia coli. Biochemistry* **31** 11231-11235 (1992)